806489

ALRAND REPORT 50A

STATISTICAL TRAINING MANUAL

Volume II

Operations Analysis Department
U. S. Navy Fleet Material Support Office
Mechanicsburg, Pennsylvania
10 November 1966

3

PREFACE

This report is based on a series of lectures on probability and statistics presented to employees of the Operations Analysis Department, U. S. Navy Fleet Material Support Office (FMSO) by Dr. Barnard H. Bissinger, Chairman of the Mathematics Department, Lebanon Valley College, Annville, Pennsylvania and Consultant to FMSO. They represent a more advanced state of development and are intended as a follow-on course to material covered in ALRAND Report 50 of 3 September 1965, "Statistical Training Manual - Volume I."

It is hoped that the manual will help other units who wish to provide training in this type of mathematics. Any corrections or remarks indicating improvement will be gratefully accepted.

Submitted by: B. H. Bissinger
B. H. BISSINGER

Approved by: J. A. Gillespie
J. A. GILLESPIE, Commander, SC, USN
Director, Operations Analysis Department

P. F. Cosgrove Jr.
P. F. COSGROVE, JR., Captain, SC, USN
Commanding Officer, U. S. Navy Fleet Material Support Office

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

(6)

TABLE OF CONTENTS (Continued)

I. CONFIDENCE

A. Floating Interval.

We concluded our previous course while we were discussing how the basics of sampling theory are used to obtain information about samples randomly drawn from a known population. Specifically we considered the mean $(\bar{x})$ or sum $S(x)$ of the elements of a sample and how other such means or sums of other random samples of the same size from the same universe might be related to this particular one. We considered the distribution of this sample statistic and declared the means to be normally distributed when the sample size was 30 or more. So we used a sample in class to learn more about samples by way of their means and also by way of their standard deviations.

The every-day practical problem requires us to use known samples and infer conclusions about the unknown population from which the sample comes; e.g., what is the population mean when the sample mean is known? Our rudiments of sampling theory will help us to make such a determination. Initially we will consider the problem of describing the population parameter from its corresponding sample statistic when the sample statistic is the mean.

The oldest method of making such an estimate was introduced by LaPlace in 1814 in dealing with the problem of inferring the value of the probability of success (p) in the binomial distribution from an observed value of the random variable x of the distribution. He regarded the

size of an interval which would include p as fixed but thought of p as a random variable. He was confident that a certain percentage of the time p would be in the interval and as a result it later became known as a confidence interval. It was not until 1927 that the correct interpretation of the interval as a random or floating interval was given by E. B. Wilson. Let us go through such an argument.

First, recall a few general facts. The sampling distribution of means is a frequency distribution of the means of all samples of a particular size each of which is drawn randomly from the same population. The mean of the sampling distribution of means tends to be very close to the population mean, although individual sample means may vary quite a bit from this value. However their variability is probably much smaller than the variability of the observations in the population. It decreases with increases in the sample size. For a large size sample the standard deviation of the values in the sample will not be very different from that of the population. Finally for many large size samples, the sampling distribution of their means is essentially a normal distribution.

So we estimate the population standard deviation σ_X by the sample standard deviation; call it s. Then we obtain an estimated standard error (deviation) of the mean by dividing s by $\sqrt{n}$, which we can call $s_{\overline{X}}$ where n is the sample size. Next, in making a guess about the population mean, we decide on what level of confidence (probability of being correct) we want. This determines for us the confidence interval or

confidence limits within which the population mean should lie. It specifies a range of values. To increase the confidence level we must make the estimate less precise. On the other hand we can be more precise if we are willing to take a bigger risk (less confidence). For example, suppose we have sampled a universe and developed a mean of 100 and a standard deviation of 7. We desire a confidence level of 95%, which means we expect the sample mean to be within our confidence interval 95 times for 100 samples. Therefore we would expect to experience sample means between 86 and 114 in all but 5% of the samples drawn. Now if we desire to be more precise, we narrow the confidence interval. If we establish the confidence interval as 93 and 107, we expect our mean to be within the confidence interval only 68% of the time. So precision is sacrificed to high level of confidence and vice versa. However, both precision and confidence level can be increased by increasing the sample size.

Now a very important point has been blithely skipped over in the last paragraph on procedure. Recall that we learned how to calculate the probability that went with a certain distance from the mean of a normal distribution to a value of the variable which had that mean. How do we suddenly slip over to using a single value of the variable and distance about it to pick up the mean? This is the approach LaPlace failed to conceive.

For example, you will recall that for a fairly normal distribution of sample means, you are fairly sure that about 68% of all possible

sample means will be within ± 1 standard errors of the mean of these means which in turn is the population mean. About 95% of all possible sample means will be within ± 2 standard errors of the population mean. Or we can say the probability of finding a sample whose mean is more than 1 standard deviation from the population mean is .32. Also, we can say the probability of the mean of the sample being more than 2 standard deviations from the mean of the population is only .05.

Consequently we can expect 95% of the time to get a sample whose mean is no farther away from the population mean than ± 2 standard errors of the sampling means. Hence the same size interval centered on every possible sample mean will pick up the population mean about 95% of the time. Therefore when it is placed on one such sample mean, we can be 95% confident of picking up the mean. Actually these last remarks constitute what we mean by 95% confidence and as such are definitions.

The above is so easy to say symbolically that the needed concept of the floating interval is often lost to the learner. For the situation as pictured below in Figure 1 we can say

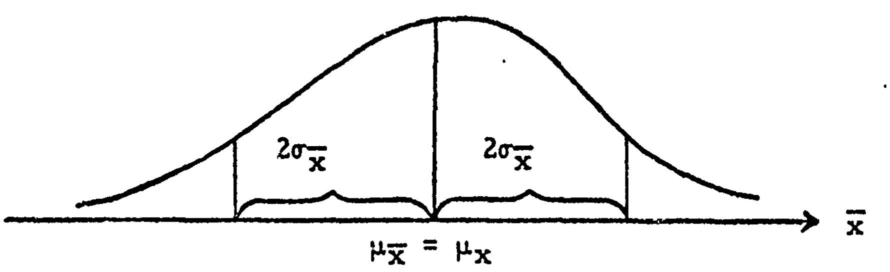


Figure 1

$$\Pr\{-2 < \frac{\bar{x} - \mu}{s/\sqrt{n}} < +2\} \doteq .95$$

which can be rewritten as

$$\Pr\{\bar{x} - 2\frac{s}{\sqrt{n}} < \mu < \bar{x} + 2\frac{s}{\sqrt{n}}\} \doteq .95$$

where n is fairly large and $\sigma_{\bar{x}}$ is estimated by $s/\sqrt{n}$. This exemplifies

$$\left(\bar{x} - 2\frac{s}{\sqrt{n}},\ \bar{x} + 2\frac{s}{\sqrt{n}}\right)$$

as an observable random interval such that the probability is .95 that it contains μ. It is a 95% confidence interval for μ and .95 is the confidence coefficient.

We have set up an estimator for a parameter by using a random interval with a specified probability of including the true value of the parameter. Such a device is called an interval estimator.

B. Floating Interval Again.

To summarize, we realize in a practical situation that we have only one sample and one mean. We have seen how all possible means behave under chance variation, but we have no way of knowing whether our single sample mean is at a point A or B or a point C, or at any other point along the $\bar{x}$ scale in Figure 2.

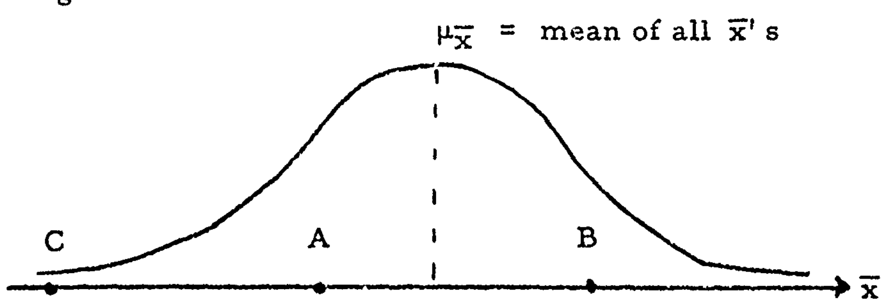


Figure 2

We have said we shall estimate the location of $\mu_{\overline{x}}$ by going out a distance d on either side of $\overline{x}$ and then claim $\mu_{\overline{x}}$ is in this interval. Now you know from the theory of the normal curve that if our claim is to be correct, the distance d will depend upon the width of the hump in the graph in Figure 3. That is, d depends on the size of $\sigma_{\overline{x}}$. If $\sigma_{\overline{x}}$ is small, the distance d does not need to be large in order to ensure that $\mu_{\overline{x}}$ is between $\overline{x} - d$ and $\overline{x} + d$. If $\sigma_{\overline{x}}$ is large, the sample means are more scattered and so a larger d will be necessary for an accurate estimate of $\mu_{\overline{x}}$.

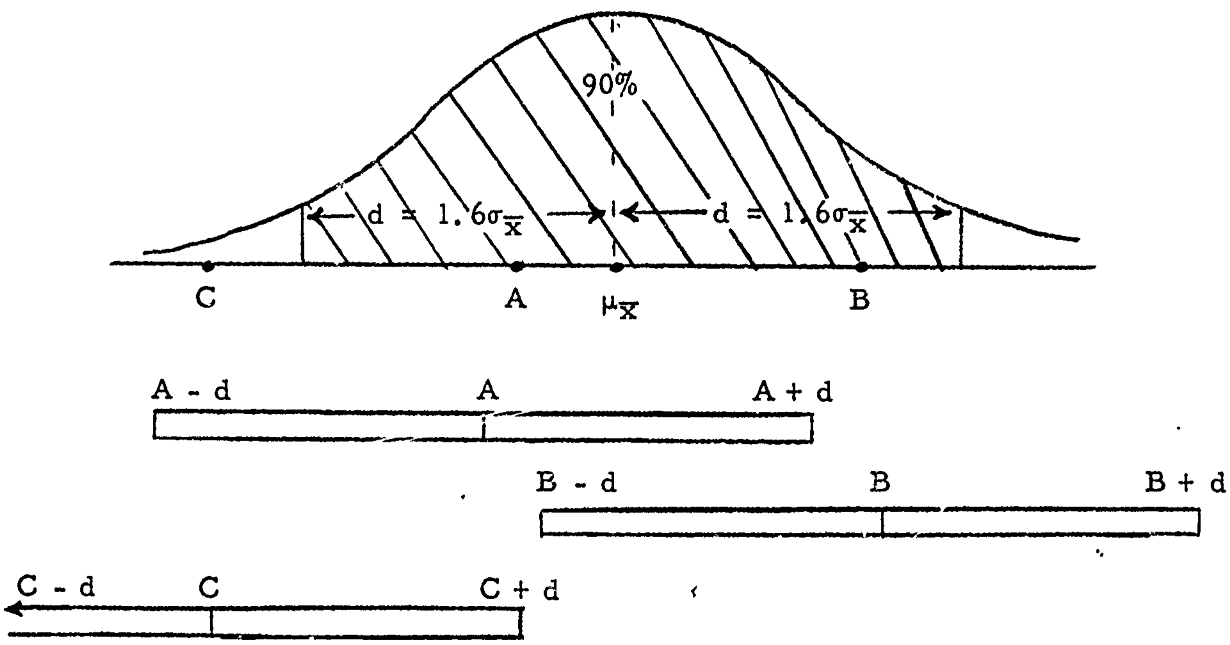
The size of $\sigma_{\overline{x}}$ measures the reliability of the mean, or the extent to which $\overline{x}$ is expected to be in error from $\mu_{\overline{x}}$ simply by chance variation. We have seen that the mean of the sample becomes more reliable as the size of the sample increases, for then $\sigma_{\overline{x}}$ decreases. So we can always rely on the sample size to pump more reliability into our estimate if time and expense permit a larger sample.

On the other hand even if $\sigma_{\overline{x}}$ is small, there can still be sample means as far away from $\mu_{\overline{x}}$ as is the point C in Figure 3. And although we take a d large enough so that the intervals $A \pm d$ and $B \pm d$ include $\mu_{\overline{x}}$, that is, so that they give a correct claim to the location of $\mu_{\overline{x}}$, the same d may not be large enough to make correct the claim that $\mu_{\overline{x}}$ is in the interval $C \pm d$. But remember that we are willing to run a specified risk of making an inaccurate claim.

Let us agree that we need to be only 90% confident that our claim about $\mu_{\overline{x}}$ is true, that is, we should expect only 9 out of 10 such claims

(12)

to be correct. This means that we shall take d large enough so that the claim will be correct for 90% of all possible sample means. Now the area table for the normal curve tells you that 90% of all the cases in a normal distribution are no more than 1.6 standard deviations from the mean of the distribution. So 90% of all sample means are within a distance of $\pm 1.6\sigma_{\overline{x}}$ of $\mu_{\overline{x}}$. Hence if we make the claim that $\mu_{\overline{x}}$ is in the interval from $\overline{x} - 1.6\sigma_{\overline{x}}$ to $\overline{x} + 1.6\sigma_{\overline{x}}$, we can be 90% certain that our claim is correct. This is true because only 10% of all possible sample means are like C, which is farther away from $\mu_{\overline{x}}$ than $1.6\sigma_{\overline{x}}$ as illustrated in Figure 3.



Three "Floating" Intervals

Figure 3

It is worthy of noting that these confidence intervals are determined fully, with exact probabilities, without assuming any a priori probability distribution for the parameter. This may seem paradoxical to you for how can we speak of the probability that a parameter lies in an interval when the parameter has no probability distribution. The answer lies in the fact that the ends of the interval vary at random in repetitions of the experiment, while the parameter point remains fixed. It' s all in the way you say it.

There is one additional point to emphasize. In our formulas on page 5, we use for the standard deviation of the base population the estimate s from the sample. Justification for this was discussed on page 2. Now from one sample to another this estimate may change a bit. Hence in Figure 3 the lengths of the three floating intervals might better be shown to be slightly different when $\sigma_{\overline{X}}$ is so obtained. However the conclusions maintain as before.

C. Project - Simulation.

1. We will select at random 10 samples of size 16 from a population with known mean and standard deviation.

2. For each sample compute $\overline{x}$ and s_X.

3. Compute an estimate of the base population mean μ at each of the confidence levels 90%, 95%, and 99% for each sample above.

4. Use the true base population standard deviation to compute the same estimates as in 3. Examine to what extent this changes the results of 3.

10

5. Note how many of these 10 estimates actually contain the true base population mean and compare this to the number we should expect from the theory of sampling. Again compare the results using the sample standard deviation with those using the true standard deviation.

To develop our solution, let us assume a frequency distribution as shown in Table I. Then we will create some device such that the numbers 0 through 10 appear in accordance with our assumed frequency. That is, 0 will appear once and there will be 48 counters marked #5. Numerous gadgets can be devised. Suppose we have a free turning gear with 200 teeth. Each of the teeth is marked with one of our numbers in a random manner so that when the gear is set in motion and then stopped any tooth has an equal chance of stopping at our reference point. We should expect to see #5 at the reference point 48 times as often as #10. With this device we can proceed with the simulation. Another commonly used device for random selection is to mark slips of paper with numbers, and put them into a hat. The hat is filled with 200 slips of paper, in this case, each marked with a number in accordance with Table I. The slips of paper (counters) are thoroughly mixed so that all counters have equal opportunity for selection. In the class we had no mechanical device, thus the hat was used.

Table I

Mark printed on counter	x	0	1	2	3	4	5	6	7	8	9	10
Number of counters	200 f[x]	1	3	10	23	39	48	39	23	10	3	1

(11)

The mean of the total counter population is 5.0 and the true standard deviation is 1.715.

Select a random sample of 16 counters by replacement. That is, draw a counter, record its number, and then replace it. Mix the counters before each selection. (The cooperative efforts of the members of the class produced the following 10 samples.)

Table II

Sample Number	0	1	2	3	4	5	6	7	8	9	10	$\overline{x}$	s_x
1			2	3	5	2	1	3				4.38	1.20
2	1	2	1	2	4	2	2	2				3.94	2.00
3				3	4	4	4	1				4.75	1.24
4	1		1	2	1	1	7	1	2			5.13	1.65
5			2	3		3	5	2	1			5.00	1.86
6	1	1	2		5	2	3	1	1			4.25	2.17
7			1		5	4	3	2	1			5.13	1.41
8	1			2	1	3	4	2	3			5.44	2.16
9	1		1	3	3	4	1	1		2		4.63	2.36
10					2	3	8	2		1		4.94	1.39

Before tabulating for each sample the additional required information let us develop it clearly for sample 1. If c represents the degree of confidence and z_c the corresponding coefficient or the number of standard

deviations from normal theory, then our three intervals are given below each followed by the words "yes" or "no," depending on whether they did or did not pick up the true mean.

$c = .90, \; z_c = 1.65, \; 4.38 \pm (1.65)(1.20)/4$

$c = .95, \; z_c = 1.96, \; 4.38 \pm (1.96)(1.20)/4$

$c = .99, \; z_c = 2.58, \; 4.38 \pm (2.58)(1.20)/4$

(3.88, 4.87) No

(3.79, 4.97) No

(3.60, 5.15) Yes

A similar calculation for each of the other nine samples yields the results in Table III.

Table III

	Confidence Interval		
Sample Nr	c = .90	c = .95	c = .99
1	3.88 - 4.87	3.79 - 4.97	3.60 - 5.15
2	3.11 - 4.76	2.96 - 4.92	2.65 - 5.23
3	4.24 - 5.26		
4	4.44 - 5.81		
5	4.13 - 5.87		
6	3.35 - 5.15		
7	4.54 - 5.71		
8	4.55 - 6.31		
9	3.65 - 5.60		
10	4.37 - 5.51		

(11)

Note that when the 90% confidence interval contained the true mean, we did not bother to calculate an interval for greater confidence as it would automatically also contain it.

In summary we can write

Table IV

Confidence in Percentage	Percentage of Times Mean Included
90	80
95	80
99	100

Remember we are straining the use of normal theory by using samples as small as 16 in size. However this strain was overcome by our taking a base population which was fairly normal itself. Also we took only 10 such samples and couldn't possibly obtain percentages of times the mean was picked up to a finer difference than 10%. Nevertheless the above project should instill into you a feeling for and a clear knowledge of the concept of a confidence interval. Ideally we want the percentage and corresponding confidence in a row to agree.

D. The Binomial Distribution.

We learned earlier that the binomial distribution

$$f_B(x) = C_x^n p^x q^{n-x}, \; x = 0, \cdots, n$$

where

$$C_x^n = \frac{n!}{x!\,(n - x)!}$$

n = sample size

p = probability of success

$q = 1 - p$

is approximately normally distributed with mean np and standard deviation $\sqrt{npq}$. As we saw in the last course, the probability of x being within a distance of $z\sqrt{npq}$ units of np is given approximately by $2F_N^*(z) - 1$, i.e.,

$$\Pr\{np - z\sqrt{npq} < x < np + z\sqrt{npq}\} \doteq 2F_N^*(z) - 1$$

For example, if $n = 400$, $p = .2$, $q = .8$, then $np = 80$, $\sqrt{npq} = 8$ and for a probability of .95 (when $z = 2$) we know that the interval (64, 96) will contain x about 95% of the time.

Now suppose we have a binomial distribution in which p is not known and from n trials we found x occurrences. We let z_α represent the coefficient and α would equal 2.5% if we use 2 standard deviations or the 95% confidence level. Then we can say

$$\Pr\{-z_\alpha < \frac{x - np}{\sqrt{npq}} < +z_\alpha\} \doteq 1 - 2\alpha$$

where $\Pr\{z > z_\alpha\} = \alpha$. The two confidence limits for p are such that

$$\frac{x - np}{\sqrt{npq}} \doteq \pm z_\alpha$$

or are the roots of

$$p^2(n^2 + nz_\alpha^2) - p(2nx + nz_\alpha^2) + x^2 = 0.$$

Solving this quadratic in p, we find these two values are

$$\frac{n}{n + z_\alpha^2}\left[\frac{x}{n} + \frac{z_\alpha^2}{2n} \pm z_\alpha \sqrt{\frac{x(n - x)}{n^3} + \frac{z_\alpha^2}{4n^2}}\right]$$

Note as the sample size n increases the above formula reduces to (we must assume x increases so that x/n doesn't fade as it is really this proportion we obtain to estimate p)

$$\left[\frac{x}{n} \pm z_\alpha \sqrt{\frac{(x/n)(1 - x/n)}{n}}\right].$$

This conforms to our normal theory which would give

$$\frac{x}{n} \pm z_\alpha \cdot \frac{\sigma_x}{n}.$$

If the population happened to be finite of size N, we would have to correct our standard deviation wherever it occurs by multiplying it by

$$\sqrt{\frac{N - n}{N - 1}}\ .$$

Consider the case when $N = 101$ and $n = 37$. Now this is sufficient to assume approximate normality (usually assumed when $n \geq 30$ and $N \geq 100$). However the correction factor becomes $\sqrt{(101 - 37)/(101 - 1)} = 0.8$. Hence we must replace the standard deviation in the above confidence interval estimate by .8 of itself. Only when N is very large compared with n is the factor nearly 1 and hence negligible.

In 1934 Clopper and Pearson in Biometrika constructed intervals of the type just discussed for p and presented graphs for 95% and 97.5%

confidence levels of p for some values of n from 10 to 1000. Instead of x, they used the sample estimate $\hat{p} = x/n$.

E. The Poisson Distribution.

A discussion, similar to that just given for the binomial distribution, can be made for the case when the base population is Poisson distributed. W. E. Ricker, following the original lines of Clopper and Pearson, presented this in 1937 in the Journal of the American Statistical Association. He gave the formula

$$x + 1.92 \pm 1.960\sqrt{x + 1.0}$$

for the 95% confidence limits of $\mu = \lambda$ for an observed value of x, while for 99% confidence he gave

$$x + 3.32 \pm 2.576\sqrt{x + 1.7}\ .$$

Actually, Professor Pearson suggested this to him via the fact that the Poisson distribution gets more and more normal as the mean λ increases so that the end-points of our random interval for a confidence of $1 - 2\alpha$ is

$$\lambda^2 - \lambda(2x + z_\alpha^2) + x^2 = 0.$$

Limiting ourselves to large values of x that might occur in a sample, we sometimes consider the result as an estimate of the mean, hence also of the variance of the assumed base Poisson distribution to which it belonged. Then the estimating random interval end-points, say for 95% confidence, are taken to be

$$x \pm 1.960\sqrt{x}.$$

By way of comparison we have for $x = 50$ the estimators by the essentially two different methods as given in Table V.

Table V

	Confidence			
	95%		99%	
Formulas Used	Lower Limit	Upper Limit	Lower Limit	Upper Limit
Pearson-Ricker	37.9	65.9	34.8	71.8
Old Method	36.1	63.9	31.8	68.2

F. Examples Using Confidence.

1. Problem 1. Suppose you know a certain part has its quarterly demands uniformly distributed over some interval of demand sizes whose smallest value is zero. You wish to estimate the upper end-point, call it a, of the interval. Now suppose you have a sample of size 20 and its mean is 3.2. What are the 90% confidence limits for a?

a. Solution. The distribution function can be written

$$f(x) = \frac{1}{a}, \quad 0 \le x \le a.$$

Now its mean and variance are easily computed to be

$$\mu = \int_0^a x \frac{1}{a} dx = \frac{a}{2}$$

$$\sigma^2 = \int_0^a x^2 \frac{1}{a} dx - \mu^2 = \frac{a^2}{12} .$$

For $1 - 2\alpha = .90$, we find $z_\alpha = 1.645$. For samples of size n, the standard deviation of the means of such samples is

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{a/\sqrt{12}}{\sqrt{n}} = \frac{a}{\sqrt{12n}}$$

Hence

$$\Pr\left\{-1.645 < \frac{\overline{x} - \frac{a}{2}}{a/\sqrt{12n}} < +1.645\right\} \doteq .9 \quad .$$

So the two values we seek for bounding a are

$$a = \frac{\overline{x}}{\frac{1}{2} \pm \frac{1.6}{\sqrt{12n}}}$$

When $n = 20$ and $\overline{x} = 3.2$, this gives the values 5.3 and 8.1. So our confidence interval estimate for a is (5.3, 8.1).

2. Problem 2. A sample of 100 stock items indicated 55% were on hand. Find 95% confidence limits for the proportion of on-hand items in the entire stock.

a. Solution. This is, like Problem 1, calling for a two-sided interval estimator. The estimate as given by the formulation on page 14 is

$$.55 \pm 1.96\sqrt{\frac{(.55)(.45)}{100}} = .55 \pm .10$$

Therefore we can be 95% confident that the true proportion lies in the interval (.45, .65).

3. Problem 3. An analysis of 40 randomly selected requisition cards revealed that 24 were from the same Navy Supply Center. Find

95% confidence limits for the actual proportion of such cards to be expected in the long run from this same center.

a. Solution. If we assume the binomial distribution with $n = 40$ and $\hat{p} = 24/40 = .6$, then the Clopper-Pearson tables give us the interval estimator (.45, .74). If instead, we assume the normal distribution and use the approximating formula, we get

$$.60 \pm 1.96\sqrt{\frac{(.6)(.4)}{40}}$$

or the confidence interval (.45, .75).

4. Problem 4. Suppose our random variable x is gamma distributed and that from a sample we find the first decile (10% cumulation) is 1.33 while the ninth decile (90% cumulation) is 5.62. Find the shape and scale of the parameter.

a. Solution. This is a deterministic problem in that the two empirical values given completely determine α and β. To see this, compute the value of the quotient of the 10 percentile value and the 90 percentile value, namely 1.33/5.62 or .236. Note in Table VI that this value of this ratio is found in the right-hand column and opposite to $\alpha = 2.5$. Now in Table VII opposite to $\alpha = 2.5$ we find x/β is 1.417 at the 10th percentile and is 6.008 at the 90th percentile. This overdeterministic situation gives the following two equations for β

$$\frac{1.33}{\beta} = 1.417 \quad \text{and} \quad \frac{5.62}{\beta} = 6.008$$

$$\beta = \frac{1.33}{1.417} = .938 \quad \beta = \frac{5.62}{6.008} = .935$$

Since each yields essentially the same value, .94, we accept it and feel some justification in the assumption of the gamma distribution. No confidence estimation was used here.

5. Problem 5. We know that a demand random variable is normally distributed $N(\mu, \sigma^2)$ and $\sigma^2 = 100$. A sample of size 25 is drawn and the observed mean $\bar{x}$ is 250. Find 95% confidence limits for the unknown population mean μ.

a. Solution. Since the sample size is close to the border value of 30 beyond which we usually assume the sample means are normally distributed, we might as well invoke the same hypothesis. Then our confidence interval becomes

$$\mu = 250 \pm \frac{1.96(10)}{\sqrt{25}} = 250 \pm 3.9$$

or (246.1, 253.9).

6. Problem 6. From a population of unknown parameter p representing a proportion having an attribute, a sample of 400 yields 320 with this attribute. Find 90% confidence limits for p, the true probability of the attribute.

a. Solution. Denoting our empirical value of p by $\hat{p}$, our 90% confidence interval is given by

$$\hat{p} \pm 1.65\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

which for $\hat{p} = .8$ and $n = 400$ becomes (.767, .833).

21

Table VI

Ratios Facilitating the Estimation of the Parameters α, β of the Gamma Distribution

α	D_1/M	D_5/M	D_9/M	D_1/D_5	D_9/D_5	D_1/D_9
-.5	Curve J-Shaped			.0348	5.960	.0058
0	Curve J-Shaped			.152	3.323	.0455
.5	*	2.366	6.252	.247	2.642	.0934
1.0	.532	1.678	3.890	.317	2.318	.137
1.5	.537	1.451	3.079	.370	2.122	.174
2.0	.551	1.337	2.661	.412	1.990	.207
2.5	.567	1.269	2.403	.447	1.893	.236
3.0	.582	1.224	2.227	.475	1.819	.261
3.5	.595	1.192	2.098	.500	1.760	.284
4.0	.608	1.168	1.999	.521	1.711	.304
4.5	.620	1.149	1.920	.539	1.671	.323
5.0	.630	1.134	1.855	.556	1.636	.340
5.5	.640	1.122	1.801	.571	1.606	.355
6.0	.649	1.112	1.755	.584	1.579	.370
6.5	.657	1.103	1.716	.596	1.556	.383
7.0	.665	1.096	1.682	.607	1.535	.396
7.5	.672	1.089	1.651	.617	1.516	.407
8.0	.679	1.084	1.624	.627	1.499	.418
8.5	.685	1.079	1.600	.635	1.483	.428
9.0	.691	1.074	1.578	.643	1.469	.438
9.5	.697	1.070	1.559	.651	1.456	.447
10.0	.702	1.067	1.541	.658	1.444	.456
11.0	.712	1.061	1.509	.671	1.423	.472
12.0	.720	1.056	1.482	.682	1.404	.486
13.0	.728	1.051	1.458	.693	1.387	.500
14.0	.736	1.048	1.438	.702	1.372	.512
15.0	.742	1.045	1.420	.711	1.359	.523
20.0	.769	1.033	1.352	.744	1.309	.569
25.0	.789	1.027	1.308	.768	1.274	.603
30.0	.804	1.022	1.277	.786	1.249	.629

*Mode to left of D_i. Where: D_i = i^{th} decile; M = mode.

Table VII

Selected Percentage Points of the Gamma Distribution: Values of x/β

Corresponding to Stated Values of $F(x)$

α \ F	.05	.10	.25	.50	.75	.90	.95
-.5	.00197	.00790	.0508	.227	.662	1.353	1.921
0	.0513	.105	.288	.693	1.386	2.303	2.996
.5	.176	.292	.606	1.183	2.054	3.126	3.907
1.0	.355	.532	.961	1.678	2.693	3.890	4.744
1.5	.573	.805	1.337	2.176	3.313	4.618	5.535
2.0	.818	1.102	1.727	2.674	3.920	5.322	6.296
2.5	1.084	1.417	2.127	3.173	4.519	6.008	7.034
3.0	1.366	1.745	2.535	3.672	5.109	6.681	7.754
3.5	1.663	2.084	2.949	4.171	5.694	7.342	8.460
4.0	1.970	2.433	3.369	4.671	6.274	7.994	9.154
4.5	2.287	2.789	3.792	5.170	6.850	8.638	9.838
5.0	2.613	3.152	4.219	5.670	7.423	9.275	10.513
5.5	2.946	3.521	4.650	6.170	7.992	9.906	11.181
6.0	3.285	3.895	5.083	6.670	8.558	10.532	11.842
6.5	3.630	4.273	5.518	7.169	9.123	11.154	12.498
7.0	3.981	4.656	5.956	7.669	9.684	11.771	13.148
7.5	4.336	5.043	6.396	8.169	10.244	12.384	13.794
8.0	4.695	5.432	6.838	8.669	10.802	12.995	14.435
8.5	5.058	5.825	7.281	9.169	11.359	13.602	15.072
9.0	5.425	6.221	7.726	9.669	11.914	14.206	15.705
9.5	5.796	6.620	8.172	10.169	12.467	14.808	16.335
10.0	6.169	7.021	8.620	10.668	13.020	15.407	16.962
11.0	6.924	7.829	9.519	11.668	14.121	16.598	18.208
12.0	7.690	8.646	10.422	12.668	15.217	17.782	19.443
13.0	8.464	9.470	11.329	13.668	16.310	18.958	20.669
14.0	9.246	10.300	12.239	14.668	17.400	20.128	21.886
15.0	10.035	11.135	13.152	15.668	18.487	21.293	23.098
20.0	14.072	15.382	17.755	20.668	23.883	27.045	29.062
25.0	18.218	19.717	22.404	25.667	29.234	32.711	34.916
30.0	22.444	24.113	27.085	30.667	34.552	38.315	40.691

G. Comparison by Two Samples.

Sometimes we want to compare the means of two samples. Really we should say we are interested in how great may be the difference between the means of their base populations. Denote these two means by μ_1 and μ_2 and also denote the sample size by n_1 and n_2, respectively, with means $\bar{x}_1$ and $\bar{x}_2$, respectively. If the samples are large, then we learned in the previous course what the variance for $\bar{x}_1 - \bar{x}_2$ is in either a finite population or an indefinitely large population in terms of the variance of each population. At that time we also remarked that the difference $\bar{x}_1 - \bar{x}_2$ is essentially normally distributed. Hence

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

is $N(0, 1)$, or

$$\Pr\left\{-z_c < \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} < + z_c\right\} \doteq c.$$

The above symbolism is a slight break with the convention of writing α for c when it is the subscript and of writing $1 - 2\alpha$ for c on the right side of the last expression. It seems more natural to write simply what we just did and realize that for, say $c = .90$, you must pick z_c so that $F(z_c) = .95$.

From our discussion in the previous course you will recall we use for $\sigma_{\bar{x}_1 - \bar{x}_2}$

$$\frac{\sigma^2_{x_1}}{n_1} + \frac{\sigma^2_{x_2}}{n_2} \quad \text{for indefinitely large populations and}$$

$$\frac{\sigma^2_{x_1}}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\sigma^2_{x_2}}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right) \text{ for finite populations.}$$

Rewriting our previous probability statement we obtain the confidence interval estimate for $\mu_1 - \mu_2$,

$$\Pr\{\bar{x}_1 - \bar{x}_2 - z_c\sigma_{\bar{x}_1 - \bar{x}_2} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_c\sigma_{\bar{x}_1 - \bar{x}_2}\} \doteq c.$$

When the σ^2_1 and σ^2_2 are not known, we may replace them by their sample estimates, s_1^2 and s_2^2, respectively.

1. Illustration. For a particular Federal Stock Number (FSN) we find that out of 580 orders in one year, the mean demand (average requisition size) is 34.4 units per order and the standard deviation is 8.83 while in the succeeding year from 786 orders the mean demand is 28.02 and the standard deviation is 8.81. What are the 95% confidence limits for the difference of the means of the two conceptually different populations?

a. Solution. For c = .95 we have $z_c = 1.96$. Therefore the limits sought are

$$(34.45 - 28.02) \pm 1.96\sqrt{\frac{(8.83)^2}{580} + \frac{(8.81)^2}{786}}$$

$$6.43 \pm 1.96 \times .84$$

$$\text{or} \quad 6.43 \pm .95$$

So we are 95% certain that if there are two different patterns of behavior for each year, the means differ by no less than 5.48 and by no more than 7.38.

If we are interested in confidence limits for the difference of two population proportions, p_1 and p_2, then really this is simply another

application of the general theory we just learned. The excuse for remarking on it lies in the simplicity of the resulting expression. Suppose $\hat{p}_1$ and $\hat{p}_2$ are the sample estimates from samples of size n_1 and n_2, respectively. Then

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{p_1 - p_2}}$$

is N(0, 1) and so

$$\Pr\left\{-z_c < \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sigma_{p_1 - p_2}} < + z_c\right\} \doteq c.$$

Now using the sample estimates for our required variances we have

$$\sigma^2_{p_1 - p_2} = \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}$$

for indefinitely large populations

or

$$= \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)$$

for large finite populations.

Rewriting our previous probability statement we obtain the confidence interval estimate for $p_1 - p_2$,

$$\Pr\{\hat{p}_1 - \hat{p}_2 - z_c\sigma_{p_1 - p_2} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_c\sigma_{p_1 - p_2}\} \doteq c.$$

2. Illustration. One FSN was ordered in 230 days out of 400 days while another was requested 200 out of 500 days. Find 95% confidence intervals for the difference between the conceptual rates of demand.

a. Solution. Assume indefinitely large populations. For $c = .95$ we have $z_c = 1.96$. Therefore the limits sought are

$$\left(\frac{230}{400} - \frac{200}{500}\right) \pm 1.96\sqrt{\frac{(230/400)(1 - 230/400)}{400} + \frac{(200/500)(1 - 200/500)}{500}}$$

$$(.575 - .400) \pm 1.96\sqrt{\frac{(.575)(.425)}{400} + \frac{(.400)(.600)}{500}}$$

or $\quad .175 \pm .065$

So for all practical purposes we might say the difference between the mean demand rates lies between .110 and .240.

II. SMALL SAMPLE THEORY

A. Some History of Research and Development.

In 1908 in the paper entitled "The Probable Error of a Mean" appearing in Biometrika, W. S. Gossett, alias "Student," wrote

> "Any experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.
>
> "Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a greater number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.
>
> "If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty: (1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.
>
> "The usual method of determining the probability that the mean of the population lies within a given distance of the mean of the sample is to assume a normal distribution about the mean of the sample with a standard deviation equal to $s/\sqrt{n}$, where s is the standard deviation of the sample, and to use the tables of the probability integral.
>
> "But as we decrease the number of experiments, the value of the standard deviation found from the sample of

experiments becomes itself subject to an increasing error, until judgments reached in this way become altogether misleading."

A few paragraphs later, Mr. Gossett goes on to say

> "Again, although it is well known that the method of using the normal curve is only trustworthy when the sample is "large," no one has yet told us very clearly where the limit between "large" and "small" is to be drawn.
>
> "The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few."

The reader must be wondering by now why we classify this concern under the heading "Small Sample Theory." Actually it's not the size of the sample that is the basic concern--it is the estimating of the base population standard deviation from the sample and this estimate goes to the lean side when the sample size is small. However, if we know the standard deviation of our base population for which we are attempting to ascertain the mean and if the base population is essentially normal, then the means of samples of any size are normally distributed and we use the z_c for our confidence c on the base population standard deviation divided by $\sqrt{n}$.

The problem arises, as Gossett said, when the base population is unknown, even though assumed normal, because then we cannot use the z_c confidence limits since σ is not known. That is, we don't know when we can use it, supposing there are such times, and further, when we can't use it, we need to know how to modify z_c to get confidence c.

The sum and substance of the mathematical problem is to find for a sample $(x_1, x_2, \cdots, x_n)$ of size n from a population $N(\mu, \sigma^2)$ the

theoretical distribution of the random variable

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

where s is the standard deviation of the sample set of numbers.

Fortunately it can be proven that this distribution function does not involve σ, the population standard deviation. Mr. Gossett first obtained the distribution of s^2 in random samples after having examined many empirical situations. He did this by using the relation connecting the first four moments of the Pearson Type III curve

$$y = A(x - \mu)^{\lambda - 1} e^{-\alpha(x-\mu)}, \quad x > \mu, \ \alpha > 0, \ \lambda > 0,$$

which generalizes the gamma distribution and hence also the chi-square distribution to be studied later. Knowledge of the first four moments of any frequency function belonging to Pearson's system is sufficient to determine that function.

Tediously, as Gossett puts it, he obtained the moments (M_i) of s^2 about its mean (since he used the bias formula he calculated the mean of s^2 to be $\mu_2(n - 1)/n$) to be, in order,

$$0, \ 2\mu_2^2(n - 1)/n^2, \quad 8\mu_2^3(n - 1)/n^3, \quad 12\mu_2^4(n - 1)(n + 3)/n^4,$$

so that

$$\beta_1 = 8/(n-1), \quad \beta_2 = 3(n + 3)/(n - 1),$$

where β_1 and β_2 are "shape predictors" not sensitive to magnitude of the data. These values satisfy the Pearson criterion

$$2\beta_2 - 3\beta_1 - 6 = 0$$

for a Type III curve. Consequently Gossett said he believed that s^2 followed the law

$$y = cx^p e^{-\gamma x}$$

where

$$\gamma = 2\frac{M_2}{M_3} = \frac{4\mu_2^2(n-1)n^3}{8n^2\mu_2^3(n-1)} = \frac{n}{2\mu_2}$$

$$p = \frac{4}{\beta_1} - 1 = \frac{n-1}{2} - 1 = \frac{n-3}{2} .$$

Consequently he got

$$f_{s^2}(x) = cx^{\frac{n-3}{2}} e^{-\frac{nx}{2\mu_2}} .$$

The distribution of s may be found since the frequency of s is that of s^2 and all we must do is to compress the base line suitably. Gossett reasoned

$$y_1 = \phi(s^2)$$

$$y_2 = \psi(s)$$

Then $\quad y_1\, d(s^2) = y_2\, d(s).$

$$\therefore \quad y_2 = 2sy_1$$

$$\therefore \quad y_2 = 2cs(s^2)^{\frac{n-3}{2}} e^{-\frac{ns^2}{2\mu_2}}$$

$$= 2cs^{n-2} e^{-\frac{ns^2}{2\mu_2}}$$

or

$$f_s(x) = Ax^{n-2} e^{-\frac{nx^2}{2\mu_2}} .$$

Next he derived the distribution of $z = x/s$, the distance of the mean of sample measurements in terms of s, for which he got

$$y(z) = \begin{cases} \frac{1}{2} \cdot \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \cdots \frac{5}{4} \cdot \frac{3}{2} (1+z^2)^{-n/2}, & n \text{ odd} \\ \frac{1}{\pi} \cdot \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \cdots \frac{4}{3} \cdot \frac{2}{1} (1+z^2)^{-n/2}, & n \text{ even} \end{cases}$$

or

$$y = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} (1+z^2)^{-n/2}$$

which has the following descriptive values:

$$\sigma = 1/\sqrt{n-3}, \quad \beta_1 = 0, \quad \mu_4 = 3/(n-3)(n-5), \quad \beta_2 = 3 + 6/(n-5).$$

And it is symmetric about zero so if we wanted to fit a normal to it, we would use the given formula for σ, ($\sigma = 1/\sqrt{n-3}$). Remember, $\Gamma(n+1) = n\Gamma(n)$ was generalized from the case when n is a positive integer and then $\Gamma(n+1) = n!$.

Now Gossett's original papers suffered from two defects:

1. As n increases the z-scale becomes very close.

2. Except in the case for which it was designed, n, the number in the sample, is not the best number under which to enter the table, but n - 1, the number of degrees of freedom, is .

So at Fisher's suggestion new tables were constructed with argument $t = z\sqrt{n'}$ where n' is now one less than the number in the sample, which Gossett temporarily called n'. So if we switch from z to the more familiar t, then we could say that the new variable and old variable are related by

$$t^2 = (n - 1)z^2$$

$$dt = \sqrt{n - 1}\, dz$$

Moreover we get to see again how when we stretch (or compress) units horizontally, we must do just the opposite vertically to preserve area. In this case the distribution of t is found from that of z since the frequency of t is equal to that of z so that all we have to do is expand the base line suitably. So we find written in many books

$$f_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi(n-1)}}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}$$

which may not be as appealing to some people as the original form of Mr. Gossett.

The new descriptive values are

$$\sigma = \sqrt{\frac{n-1}{n-3}}, \quad \beta_1 = 0, \quad \mu_4 = 3(n-1)^2/(n-3)(n-5),$$

$$\beta_2 = 3 + 6/(n-5).$$

The parameter $n - 1$ is called the degrees of freedom. For small n this t-distribution differs considerably from the unit normal distributions which it approaches as n increases without limit. In Figure 4 the graph for $n = 4$ is compared with that of the limiting normal. Here it can be seen that the probability of a large deviation from the mean is much larger in the t-distribution than in the normal case.

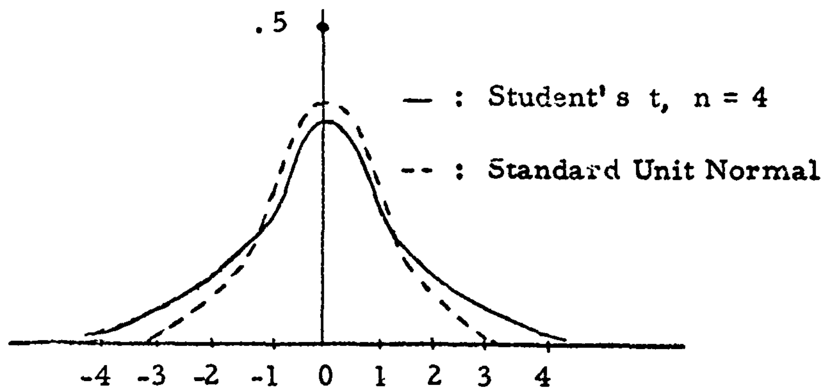


Figure 4

To see that the normal distribution is the limiting distribution we write $f_n(t)$ as

$$\left[\frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{\sqrt{2}}{\sqrt{n}}\right]\frac{1}{\sqrt{2\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

The factor in brackets can be shown to approach unity and for every fixed t,

$$-\frac{n+1}{2}\log\left(1+\frac{t^2}{n}\right) \to -\frac{t^2}{2}.$$

Hence

$$f_n(t) \to \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}.$$

All of the above remarks can be gleaned from Table VIII of values that gives t_c for confidence c. Consistent with our notation heretofore let us write

$$f_{t,\,n-1}(t)$$

for " the probability density function of the random variable known as

Table VIII. Values of t_c for Centered Confidence Interval

Sample Size n	Degrees of Freedom n - 1	c .99	c .95	c .90
2	1	63.657	12.706	6.314
3	2	9.925	4.303	2.920
4	3	5.841	3.182	2.353
5	4	4.604	2.776	2.132
6	5	4.032	2.571	2.015
7	6	3.707	2.447	1.943
8	7	3.499	2.365	1.895
9	8	3.355	2.306	1.860
10	9	3.250	2.262	1.833
11	10	3.169	2.228	1.812
12	11	3.106	2.201	1.796
13	12	3.055	2.179	1.782
14	13	3.012	2.160	1.771
15	14	2.977	2.145	1.761
16	15	2.947	2.131	1.753
17	16	2.921	2.120	1.746
18	17	2.898	2.110	1.740
19	18	2.878	2.101	1.734
20	19	2.861	2.093	1.729
21	20	2.845	2.086	1.725
22	21	2.831	2.080	1.721
23	22	2.819	2.074	1.717
24	23	2.807	2.069	1.714
25	24	2.797	2.064	1.711
26	25	2.787	2.060	1.708
27	26	2.779	2.056	1.706
28	27	2.771	2.052	1.703
29	28	2.763	2.048	1.701
30	29	2.756	2.045	1.699
31	30	2.750	2.042	1.697
41	40	2.705	2.021	1.684
61	60	2.660	2.000	1.671
121	120	2.617	1.980	1.658
∞ ← z_c →	∞	2.576	1.960	1.645

Student's t - when the sample size is n or the number of degrees of freedom is n - 1."

It is apparent from Table VIII that s underestimates σ on the average for a fixed n. For any given confidence as n decreases to zero, the confidence coefficient t increases. On the other hand for large n we see that t_c is practically z_c and that in the limit this equality exists. Also it is to be noted that t_c settles down faster for larger values of c. For example, when n goes from 10 to 11, t_c changes by only .02 for c = .90 while it changes by .1 for c = .99.

Gossett, in concluding remarks, expressed belief that if the base population distribution is not normal and if, consequently, the mean and standard deviation of a sample have greater variability, still they will tend to counteract each other, a mean deviating more from the general mean tending to be divided by a larger standard deviation. Experience in subsequent years showed him correct for small samples of size less than 30 from populations sufficiently nearly normal.

B. Using the t-distribution.

So, if we want to estimate the mean μ of a base population by using a sample of size n whose mean is $\bar{x}$ and whose standard deviation is s, we simply decide on the desired confidence c, then look up t_c for n - 1 degrees of freedom. It follows that we can say

$$Pr\left\{-t_c < \frac{\bar{x}-\mu}{s/\sqrt{n}} < +t_c\right\} \doteq c$$

or

$$\Pr\left\{\bar{x} - t_c \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_c \frac{s}{\sqrt{n}}\right\} \doteq c.$$

Once again we have a random interval

$$\left(\bar{x} - t_c \frac{s}{\sqrt{n}},\ \bar{x} + t_c \frac{s}{\sqrt{n}}\right)$$

which 100c% of the time should include μ.

1. Illustration. A set of 11 requisitions for a particular stock number has a mean $\bar{x} = 4$ (average requisition size) and a standard deviation $s = .6$. What are the 95% confidence limits for the true mean (μ) or the average requisition size for all requisitions for this stock number?

a. Argument. Since $n - 1 = 10$ = degrees of freedom and $c = .95$, we find from Table VIII that $t_c = 2.228$. Therefore the 95% confidence limits for μ are

$$4 \pm 2.228\left(\frac{.6}{\sqrt{11}}\right) = 4 \pm .4$$

or the estimating interval is (3.6, 4.4).

Suppose $\bar{x}_1$ and $\bar{x}_2$ are the means of two samples $\{x_{1i}\}$ and $\{x_{2j}\}$ of sizes n_1 and n_2, respectively, from the same base population. Then

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$$

will be normally distributed N(0, 1). Now we must estimate σ from our sample data. You may recall that when we pooled two sets of data that had the same mean we found the pooled variance to be the weighted

arithmetic mean of the two variances, namely

$$s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2).$$

Consistent with this, if we let

$$S_1 = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2, \quad S_2 = \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2,$$

we estimate σ by

$$s = \sqrt{\frac{S_1 + S_2}{n_1 + n_2 - 2}} .$$

When this sample standard deviation is used in place of σ, then

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is distributed as "Student's t" with $n_1 + n_2 - 2$ degrees of freedom.

Therefore we can say

$$\Pr\left\{ -t_c < \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < + t_c \right\} \doteq c$$

or the 100c% confidence interval estimate of $\mu_1 - \mu_2$ is

$$\left(\bar{x}_1 - \bar{x}_2 - t_c s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \ \bar{x}_1 - \bar{x}_2 + t_c s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right).$$

2. Illustration. Suppose two sets of quarterly demand observations are

D_1	D_2
16	17
16	27
20	18

16	25
20	27
17	29
15	27
21	23
	17

Estimate the difference $\mu_1 - \mu_2$ on the assumption they each came from a different base population.

a. Argument. First we calculate

$$n_1 = 8 \qquad n_2 = 9$$

$$\bar{x}_1 = \bar{D}_1 = 17.62 \qquad \bar{x}_2 = \bar{D}_2 = 23.33$$

$$S_1 = 37.88 \qquad S_2 = 184.00$$

$$S_1 + S_2 = 221.88$$

$$\bar{x}_1 - \bar{x}_2 = 5.71.$$

Then the estimated variance of the difference between the means is given by

$$\frac{s^2(n_1 + n_2)}{n_1 n_2} = \frac{(S_1 + S_2)(n_1 + n_2)}{(n_1 + n_2 - 2)(n_1 n_2)} = \frac{(221.88)(17)}{(15)(9)(8)}$$

$$= 3.50$$

and the estimated standard deviation is 1.87. Hence for degrees of freedom $(n_1 + n_2 - 2) = 15$, we have for, say 95% confidence, the interval estimate for $\mu_1 - \mu_2$ of

$$5.71 \pm 2.131(1.87) = 5.71 \pm 3.98$$

or

$$(1.73, 9.69).$$

Another conclusion of a confidence type can be drawn. We are 95% confident that the means of the two assumed base populations are different since the lower end-point of our estimating interval is positive. If the value zero had been included in our confidence interval, then equality of the means would not be rarer than 5% of the time due to chance.

3. Illustration. Here are two sets of demands that certainly appear to be alike. It must be remembered that we test by virtue of the quotient of our sample means' difference and a pooled estimate of standard deviation.

D_1	D_2
79.98	80.02
80.04	79.94
80.02	79.98
80.04	79.97
80.03	79.97
80.03	80.03
80.04	79.95
79.97	79.97
80.05	
80.03	
80.02	
80.00	
80.02	

a. Argument. We find $\overline{D}_1 = 80.02$, $\overline{D}_2 = 79.98$, $n_1 = 13$, $n_2 = 8$, $s_1^2 = .000574$, $s_2^2 = .000984$. Therefore our estimate of σ is

$$s = \sqrt{\frac{12(.000574) + 7(.000984)}{19}} = \sqrt{.000725} = .0269$$

Now for $c = .95$ and degrees of freedom = 19, we find from Table VIII that $t_c = 2.093$. So our error term becomes

$$t_c s \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 2.093 \times .0269 \times \sqrt{\frac{13 + 8}{13 \times 8}} = .025$$

and we find that with 95% confidence these two sets of demands come from separate populations the difference of whose means is $.04 \pm .025$. Or we can say we are 95% confident the true difference of the base population means lies in the interval $(.015, .065)$.

Again we can also be confident that the two base populations have different means.

C. Chi-square.

Of interest is the sample random variable

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{\sigma^2}$$

which can be written equivalently as

$$\frac{(n - 1)s^2}{\sigma^2}.$$

When the samples of size n are drawn from a normal distribution with variance σ^2, this new random variable has its density function given by

$$\frac{1}{2^{\left(\frac{n-1}{2}\right)}\Gamma\left(\frac{n-1}{2}\right)}\left[\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\left(\frac{n-1}{2}\right)-1} e^{-\frac{(n-1)s^2}{(2\sigma^2)}}\right].$$

Karl Pearson may have used an awkward symbol to replace this variable, but he wanted to "characterize a sum of squares." So he picked the Greek for "ch" which is χ and then put a 2 on it in exponential position. He called the symbol χ^2, "chi-square," and wrote the above probability density as

$$f(\chi^2) = \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}(\chi^2)^{\frac{n-3}{2}}e^{-\frac{\chi^2}{2}}$$

We say this is "χ^2 with n - 1 degrees of freedom" and that

$$\left(\frac{(n-1)s^2}{\chi^2_{UC}}, \frac{(n-1)s^2}{\chi^2_{LC}}\right)$$

is a 100c% confidence interval for σ^2 while

$$\left(\frac{\sqrt{n-1}\,s}{\sqrt{\chi^2_{UC}}}, \frac{\sqrt{n-1}\,s}{\sqrt{\chi^2_{LC}}}\right)$$

is a 100c% confidence interval for σ. In keeping with our earlier method of notation we will write

$$f_{ch^2, n-1}(\chi^2) \text{ for } f(\chi^2) .$$

Looking back to page 90 of ALRAND Report 50, Volume I, we see this is simply Γ with $\lambda = 1/2$ and $k = (n-1)/2$.

1. Illustration. The standard deviation of a random sample of 16 requisitions for an item is $9.6/\sqrt{15}$ based on units per requisition. Assuming the requisition size in units per requisition is normally distributed, find 95% confidence limits for the standard deviation of the historical

requisition size for this item and also for the size of requisitions to be experienced in the future.

a. Argument. The degrees of freedom are n - 1 = 15. For c = .95, we find in Table IX

$$\chi^2_{.975} = 27.5, \quad \chi^2_{.025} = 6.26.$$

Hence

$$\sqrt{\chi^2_{.975}} = 5.24, \quad \sqrt{\chi^2_{.025}} = 2.50$$

and our confidence interval estimate is

$$\left(\frac{\frac{9.6}{\sqrt{15}}\sqrt{15}}{5.24}, \frac{\frac{9.6}{\sqrt{15}}\sqrt{15}}{2.50} \right)$$

or

(1.83, 3.84), roughly (2, 4).

A few additional remarks are in order about this distribution function. The graphs of a few look like Figure 5.

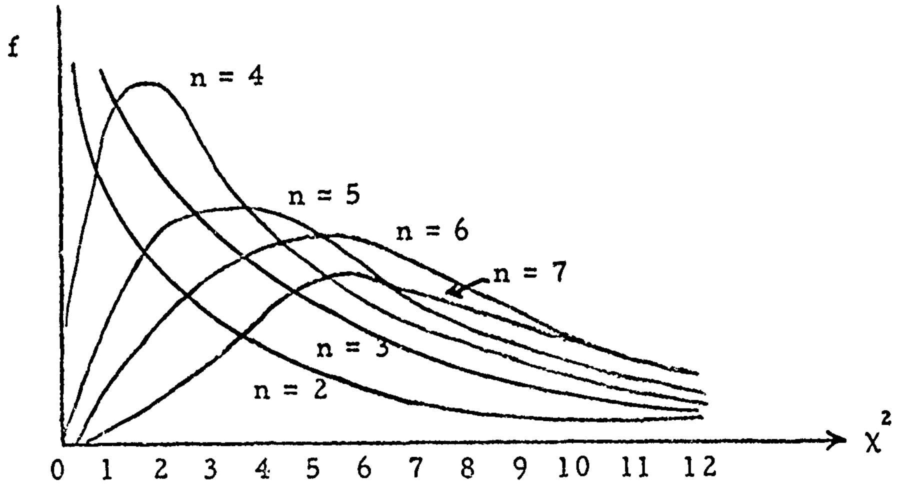


Figure 5

Table IX. Values of the Cumulative χ^2 Distribution

Sample Size n	Degrees of Freedom n - 1	Cumulative Probability .010	.025	.05	.10	.90	.95	.975	.990
2	1	.0002	.001	.0039	.0158	2.71	3.84	5.02	6.63
3	2	.0201	.0506	.103	.211	4.61	5.99	7.38	9.21
4	3	.115	.216	.352	.584	6.25	7.81	9.35	11.34
5	4	.297	.484	.711	1.064	7.78	9.49	11.14	13.28
6	5	.554	.831	1.145	1.61	9.24	11.07	12.83	15.09
7	6	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81
8	7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
9	8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
10	9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
11	10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
12	11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
13	12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
14	13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
15	14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
16	15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
17	16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
18	17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
19	18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
20	19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19
21	20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
22	21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93
23	22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29
24	23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64
25	24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98
26	25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31
27	26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64
28	27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96
29	28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
30	29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59
31	30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
41	40	22.16	24.43	26.51	29.05	51.80	55.76	59.34	63.69
51	50	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
61	60	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
71	70	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4
81	80	53.44	57.15	60.39	64.28	96.58	101.9	106.6	112.3
91	90	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1
101	100	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8

When the degrees of freedom, n', is greater than 2, the mean value is n' and the variance is $2n'$ while the mode is at $n' - 2$.

Next, to make it obvious this new distribution belongs to small sample theory, we note, when $(x_1, x_2, \cdots, x_n)$ is a random sample from $N(\mu, \sigma^2)$, that

$$\frac{\dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\dfrac{\Sigma (x_i - \bar{x})^2}{\sigma^2 (n - 1)}}} \quad \text{boils down to} \quad \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

our "Student-t" variable. Note the denominator in the first expression is the square root of our χ^2 divided by $(n - 1)$ which is equivalent to the square root of (s^2 / σ^2). Right here you might expect that χ^2 becomes normal as n gets bigger. Also the choice of the concept of degrees of freedom becomes more meaningful when we see we are really referring to the number of random variables independently chosen from the normal distribution. In symbols we have found

$$t = \frac{x}{\sqrt{\chi^2 / n}}$$

if x comes from $N(0, 1)$.

Student's t, therefore, affords the solution to a variety of problems beyond that for which it was originally intended because it is applicable to all cases which can be reduced to a comparison of the deviation of a normal variate with an independently distributed estimate of its standard deviation, derived from the sums of squares of homogeneous normal deviations either from the true mean of the distribution or from the means of samples.

We can distinguish, by virtue of degrees of freedom, by saying that for a random sample $(x_1, x_2, \cdots, x_n)$ from $N(\mu, \sigma^2)$ both of the following random variables are distributed like chi-square and further

$$\sum_1^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \text{ has } n \text{ degrees of freedom}$$

while

$$\sum_1^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \text{ has } (n - 1) \text{ degrees of freedom.}$$

There are many applications for which we need a transformed version of χ^2. Suppose our random variables are $x_1, x_2, \cdots, x_n$, each being $N(0, 1)$. Then

Variable = x	Frequency Function with n d. f.
$\sum_1^n x_i^2$	$f_{ch;n}(x)$
$\frac{1}{n} \sum_1^n x_i^2$	$nf_{ch;n}(nx)$
$\sqrt{\sum_1^n x_i^2}$	$2xf_{ch;n}(x^2)$
$\sqrt{\frac{1}{n} \sum_1^n x_i^2}$	$2nxf_{ch;n}(nx^2)$.

A fine algebraic property of χ^2 is that if χ_1^2 is of n_1 degrees of freedom, and similarly for χ_2^2, then $\chi_1^2 + \chi_2^2$ is χ^2 with $n_1 + n_2$ degrees of freedom. This reproductive property is shared by the binomial, Poisson, and normal distributions.

2. **Illustration.** Suppose you are interested in only 90% confidence in estimating the population variance σ^2 from a sample of size 10 with $s^2 = 195$. From Table IX we find

$$\chi^2_{.05} = 3.33, \quad \chi^2_{.95} = 16.92.$$

Then the interval estimate is given by

$$\Pr\{3.33 < \frac{9(195)}{\sigma^2} < 16.92\} = .90.$$

Thus the interval estimate for σ^2 is

$$\frac{9(195)}{16.92} < \sigma^2 < \frac{9(195)}{3.33}$$

or

$$103 < \sigma^2 < 527.$$

D. Fisher's Z and Snedecor's F Distributions.

In 1924 Fisher concerned himself with the distribution of quotients of sums of squares of normally distributed random variables. He called

$$\frac{\chi_1^2/n_1}{\chi_2^2/n_2}$$

e^{2Z} and found the distribution of Z. As mysteriously complicated as this appears at first, so is the reason for it that simple. He wanted to devise a testing function for the difference between two variances, s_1^2 and s_2^2, derived from two samples from normal distributions. If we went about this as we have with other statistics, we would have considered how often $s_1 - s_2$ would exceed its observed value. Of course our testing statistic

would have to have σ_1 and σ_2 in it. The only way to get rid of them would be to replace them by s_1 and s_2, respectively. But remember how we had to get a new distribution when we similarly changed the test statistic

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \text{to} \quad \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

for small samples. Here we would be trying to revamp something like

$$\frac{(s_1 - s_2) - (\sigma_1 - \sigma_2)}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \quad \text{into} \quad \frac{(s_1 - s_2) - (\sigma_1 - \sigma_2)}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

Fisher said the only exact treatment can come from eliminating the unknown σ_1 and σ_2 from the distribution by replacing the distribution of s_i by that of $\ln s_i$, $i = 1, 2$. In this way you will note our interest goes from

$$s_1 - s_2 \quad \text{to} \quad \ln s_1 - \ln s_2 \quad \text{to} \quad \ln \frac{s_1}{s_2} .$$

Moreover, whereas the sampling errors in s_i are proportional to σ_i, the sampling errors of $\ln s_1$ depend only on the size of the sample from which s_1 was calculated.

In 1934 Snedecor transformed the variable to e^{2Z} and out of honor to Fisher wrote F for e^{2Z}. He gave the probability element to be

$$f_S(F) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)\left(\frac{n_1}{n_2} F\right)^{\frac{n_1}{2} - 1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\left(1 + \frac{n_1}{n_2} F\right)^{\frac{n_1 + n_2}{2}}} \, d\left(\frac{n_1}{n_2}\right)F , \quad F > 0,$$

where we have n_1 and n_2 degrees of freedom.

This is a highly tabulated function for which Tables X and XI give values of F for 90% and 95% confidence, respectively. The following characteristics apply:

$$\text{Mean} = \mu_F = \frac{n_2}{n_2 - 2}, \quad n_2 > 2$$

$$\text{Variance} = \sigma_F^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4$$

$$\text{Mode} = \frac{n_1 - 2}{n_1} \cdot \frac{n_2}{n_2 + 2}, \quad n_1 > 2.$$

Using our definition of χ^2 for samples, we see that if $n_1 + 1$ and $n_2 + 1$ are the sample sizes (then n_1 and n_2 are the degrees of freedom),

$$\frac{\chi_1^2/n_1}{\chi_2^2/n_2} = \frac{\dfrac{n_1 s_1^2}{n_1 \sigma_1^2}}{\dfrac{n_2 s_2^2}{n_2 \sigma_2^2}} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = F.$$

Just as in the nonsymmetric case of χ^2, we here will designate, in contrast to the symmetric cases of the normal z and Student t, our lower and upper confidence limits multipliers by F_{LC} and F_{UC}, respectively.

What is tabulated is the ratio of the sample variances of different sizes from a standard unit normal distribution. Like the t-distribution, it is independent of population variance if both samples are drawn from the same population, i.e.,

$$F = \frac{s_1^2/\sigma^2}{s_2^2/\sigma^2} = \frac{s_1^2}{s_2^2}.$$

Our general probability statement, therefore, for any two different normal populations and for any two different size samples with variances s_1^2 and s_2^2 is

$$\Pr\left\{F_{LC} < \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} < F_{UC}\right\} = c$$

or

$$\Pr\left\{\frac{1}{F_{UC}}\frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{F_{LC}}\frac{s_1^2}{s_2^2}\right\} = c$$

This gives a probabilistic hold on relative precisions, if you will, from two samples and thereby on the variance ratio σ_1^2/σ_2^2.

The tables are set up for ratio values greater than unity, that is, for a larger variance in the numerator, i.e., for $F > 1$. Consequently the lower confidence limit, F_{LC}, for a fixed confidence cannot be directly read from the table. However it can be found by using the table.

Think of the ratio $F = s_1^2/s_2^2$ as in Figure 6.

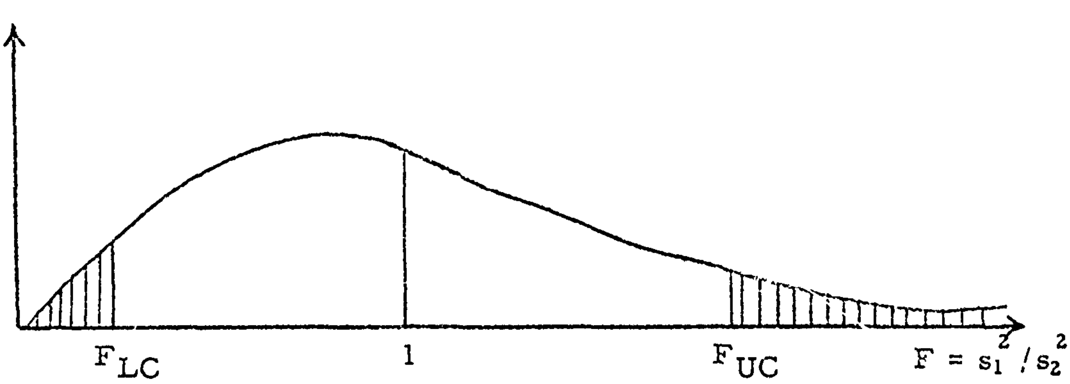


Figure 6

Since the identification of which sample's variance is in the numerator is arbitrary, we see that

$$\Pr\{F < 1\} = \Pr\{F > 1\} = 1/2.$$

Suppose we wish to find for confidence c the F_{UC} and F_{LC} when the numerator sample has n_1 degrees of freedom and the denominator sample has n_2 degrees of freedom.

Since $\Pr\{F > F'\} = \Pr\{F < 1/F'\}$,

but also $\Pr\{F < 1/F'\} = \Pr\{1/F > F'\}$,

we have $F_{LC(n_1, n_2)} = 1/F_{UC(n_2, n_1)}$.

where F' is a specific value of F.

1. Illustration. Suppose we have a sample of size 8 with variance 7.14 and a sample of size 10 with variance 3.21. Find a 90% confidence interval estimate for the quotient of the populations' variances.

a. Argument. The larger variance goes into the numerator of F, so $n_1 = 7$ and $n_2 = 9$. Then

$$F = 7.14/3.21 = 2.22.$$

From the 95% cumulative Table XI we see that

$$F_{UC(7,9)} = 3.29 \text{ while } F_{LC(7,9)} = 1/F_{UC(9,7)} = 1/3.68$$

$$\therefore \quad \Pr\left\{\frac{1}{3.68} < \frac{2.22}{\sigma_1^2/\sigma_2^2} < 3.29\right\} \doteq .9$$

$$\text{or} \quad \Pr\left\{\frac{1}{3.29}(2.22) < \frac{\sigma_1^2}{\sigma_2^2} < 3.68(2.22)\right\} \doteq .9$$

So a 90% confidence interval estimate of the quotient of the population variances is (.67, 8.18).

2. Illustration. Two different samples of size 25 each have variances 1.04 and .51. What can be said about the population variances?

a. Argument. By an argument similar to that in the previous illustration, we find for $n_1 = n_2 = 24$ and for $c = .90$ that $F_{UC} = 1.98 = 1/F_{LC}$

$$\therefore \quad \Pr\left\{\frac{1}{1.98} < \frac{1.04/.51}{\sigma_1^2/\sigma_2^2} < 1.98\right\} \doteq .9$$

$$\text{or} \quad \Pr\left\{\frac{1.04}{.51(1.98)} < \sigma_1^2/\sigma_2^2 < \frac{(1.04)(1.98)}{.51}\right\} \doteq .9$$

So (1.03, 4.04) is a 90% confidence interval estimate of the quotient σ_1^2/σ_2^2.

It should be noted that this F-distribution includes the normal distribution, the χ^2-distribution and Student's t-distribution as special cases per the following:

(1) $n_2 = \infty \rightarrow F = \chi^2/n_1$.

(2) $n_1 = \infty \rightarrow 1/F = \chi^2/n_2$.

(3) $n_1 = 1 \rightarrow \sqrt{F} = t$.

(4) $\left.\begin{array}{l} n_1 = 1 \\ n_2 = \infty \end{array}\right\} \rightarrow F = z^*$.

3. There is a very important point to make regarding the need for an F-distribution analysis before making a t-distribution analysis on the difference between two sample means such as was done in the illustration

in paragraph B3, page 38. Therein it was assumed the variances of the base populations were equal and/or both populations were the same. We pooled the two sample variances to get a good estimate of this variance which the application of Student's t needed and assumed was the same for the two base populations.

Therefore it falls upon us, prior to a Student's t analysis on the difference of the sample means, to determine whether the sample variances are enough alike to support the assumption that they are independent estimates of a common population variance. So the F-distribution should enter the scene first.

4. Finally, and in contrast, for large samples we should remember that the two sample standard deviations are analyzed by considering the variance of the actual random variable difference of the two standard deviations and that we do this by using the variance of the distribution of sample standard deviations which is $\sigma^2/2n$.

a. Illustration. Two independent samples of sizes 744 and 22 have standard deviations 1.6 and 2.1, respectively. Compare the samples with regard to the possibility of their coming from a common population. For example, such a problem might arise in dealing with superceding items. The original FSN supported a known population with known demands over time. We have collected demands on the superseding FSN for a period of time, and now we would like to know if it is supporting the same population as the superseded item.

(1) Argument. Assuming it is, we estimate the variance by pooling to be

$$s^2 = \frac{743(1.6)^2 + 21(2.1)^2}{744 + 22 - 2} \doteq 2.61$$

Therefore

$$\sigma^2_{s_1 - s_2} = \sigma^2_{s_1} + \sigma^2_{s_2} \doteq \frac{2.61}{2(744)} + \frac{2.61}{2(22)}$$

$$\doteq .0611$$

$$\sigma_{s_1 - s_2} \doteq .248 \text{ say } .25 .$$

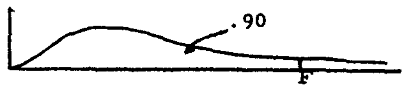
So the calculated standard deviation of the random difference between standard deviations of two such sized samples is 0.25. Thus

$$\Pr\left\{-z_c < \frac{(s_1 - s_2) - \mu_{s_1 - s_2}}{\sigma_{s_1 - s_2}} = \frac{s_1 - s_2}{0.25} < z_c\right\} \doteq c.$$

But $(s_1 - s_2)/0.25$ is $N(0,1)$, and so our particular difference, $2.1 - 1.6 = .5$ and thus two standard deviations would cover the random difference between the sample deviations so we would assume a common population.

(61)

Table X. .90 Percentile = F



$F_{s;n_1,n_2}(F) = .90$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.50	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	[illegible]	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Table XI. .95 Percentile = F

$$F_{s;n_1,n_2}(F) = .95$$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

III. TOLERANCE

A. Floating Interval.

Heretofore we have studied the theory and method of estimating magnitudes of population characteristics by intervals, namely confidence intervals. Now we wish to speak briefly of another type of interval estimate which is used when you want, so to speak, to cover a range of values and not just a single value. In particular, it is frequently desirable to make an estimate which, with certain confidence as we have used the concept, contains nearly all of the population values. There are times when you and I would like to know within what limits a certain percentage, say 99%, of the base population lies.

Obviously if we knew the mean μ and standard deviation σ of the base population and also if it was normally distributed, then $(\mu - 3\sigma, \mu + 3\sigma)$ would be a satisfactory interval. In lieu of such base population knowledge, we can use the sample mean $\overline{x}$, the sample standard deviation s, and then we can pick a k_c such that $\overline{x} - k_c s$ and $\overline{x} + k_c s$ would include 99.7% of the base population with level of confidence c. The choice of k_c depends as much on our further assumptions about the type of base population distribution as on the size of the sample.

The end points of such "floating" intervals are called statistical tolerance limits and the interval itself a statistical tolerance interval. Obviously, as the sample size increases, these intervals tend to a fixed size which depends on the percentage of base population you wish to pick up.

(60)

In contrast, confidence intervals decrease in width to zero as the sample size increases. Though both types tend to vary less in both position and width among each other for a fixed sample size, the confidence interval pinches in on the true value of the population parameter while the statistical tolerance interval tends to a fixed size since it gives limits within which an expected proportion of the population lies with some confidence.

So we see a tolerance multiplier k which depends on n, P, and c is such that we can be 100c% confident that a proportion P of the population lies between $\bar{x}$ - ks and $\bar{x}$ + ks. Now there are tables that provide us with values of k if a normal distribution can be assumed and there are tables for when it can't be assumed. Historically the latter came first from the work of S. S. Wilks and we will discuss them later in the course. For the present we will restrict ourselves to the assumption of normality and use values of k from known tables with the experimental data in an earlier illustration.

B. An Illustration.

From the data in the Project - Simulation, pages 8 to 12, tolerance intervals for each of the three different confidences .90, .95 and .99 can be calculated and then the percentage of base population they pick up can be given. For n = 16, we must modify the t-distribution coefficients 1.753, 2.131, 2.947 to the values in Table XII. The values in Table XII were taken from much more extensive tables whose reproduction here is not warranted for our immediate purpose. Further we

(66)

Table XII

c \ P	.90	.95	.99
.90	2.246	2.676	3.514
.95	2.437	2.903	3.812
.99	2.872	3.421	4.492

will restrict ourselves in the examination of the data to 90% coverage, i.e., P = .90, and to c = .90. The calculations and results are given in Table XIII, using k = 2.246.

Table XIII

Sample Number	$\bar{x}$	s	$(\bar{x} - 2.246s,\ \bar{x} + 2.246s)$	Empirical Proportion	Satisfactory Coverage
1	4.375	1.204	(1.671, 7.079) → (2, 7)	91%	Yes
2	3.938	2.00	(0, 8.430) → (0, 8)	98%	Yes
3	4.750	1.24	(1.965, 7.535) → (2, 7)	91%	Yes
4	5.125	1.65	(1.419, 8.831) → (2, 8)	96%	Yes
5	5.000	1.86	(.822, 9.178) → (1, 9)	99%	Yes
6	4.250	2.17	(0, 9.123) → (0, 9)	99.5%	Yes
7	5.125	1.41	(1.958, 8.292) → (2, 8)	96%	Yes
8	5.437	2.16	(.113, 10.761) → (1, 10)	99.5%	Yes
9	4.625	2.36	(.676, 9.926) → (1, 9)	99%	Yes
10	4.938	1.39	(1.817, 8.059) → (2, 8)	96%	Yes

(67)

On the average we should get 90% of the tolerance intervals covering 90% of the population. We did better since all of them covered at least 90%. Hence we acquire some reassurance in this illustration for saying we are 90% confident in any one sample estimate.

There are tables available that give a different value of k such that we can be 100c% confident that a proportion P of the base population will lie above (below) $\overline{x} - ks (\overline{x} + ks)$. And, as we stated earlier, we have tables for all such cases when the assumption of normality is dropped.

IV. SIGNIFICANCE

A. Significance Testing.

Suppose a person has obtained a set of observations of some process and has computed an average. He wants to know whether his statistic represents all observations for the process. If his experience is such that intuitively he believes the sample average is close to the average of the base population, he will make the statement that they are about equal. This is his hypothesis. To satisfy himself that his hypothesis is reasonable he calculates the probability that his sample could have occurred. The significance test does not tell him whether his hypothesis is right or wrong. But by choosing the proper level of significance he knows, according to probability theory, that he will seldom reject a true hypothesis, and thus he can proceed as though his assumption was fact.

A significance test involves a random sample and a probabilistic computation which decides whether or not the sample could have reasonably come from an assumed distribution. Acceptance of the assumed distribution for parenthood comes when the observed sample result is no less probable than some predetermined small probability like .10, .05, or .01. This degree of rareness due to chance is called the significance level. If the result from the sample is less probable due to chance than this, we say the result is statistically significant and we mean significant of other than chance. The region of values where probability of occurrence is greater than this is called the acceptance region while the complementary region

is called the critical or rejection region. The latter words are descriptive of the decision to reject the parenthood of the assumed base population. This is commonly called rejecting the Null Hypothesis which assumed no difference between assumed base population and the sample other than what chance allowed.

Actually this amounts to saying whether the computed confidence interval does or does not include the corresponding parameter of the base population. At the outset it appears that a confidence interval approach to making such a decision has the advantage of giving some idea of how large the difference between statistic and parameter is likely to be while a test of significance gives a cut and dried yes or no.

When we reject the Null Hypothesis, i. e., decide the discrepancy between the statistic and parameter is too rare to be due to chance, we are said to be invoking the principle of advocatus diaboli or The Devil's Advocate. This derives from the characterization of that diabolical fellow to make adverse criticism of what was deemed good.

Another word ought to be said here about the dependency of acceptance or rejection on the particular characteristic and its distribution function. We accept or reject the sample-parent association through such a device. Hence, as we will show later, it is possible, given a fixed sample and an assumed base population, to have two tests based on different statistics, one accepting and one rejecting.

Significance tests use, in general, critical regions in one of the three ways illustrated in Figure 7 for the particular value of 5% where

we have two one-tailed tests and a two-tailed test.

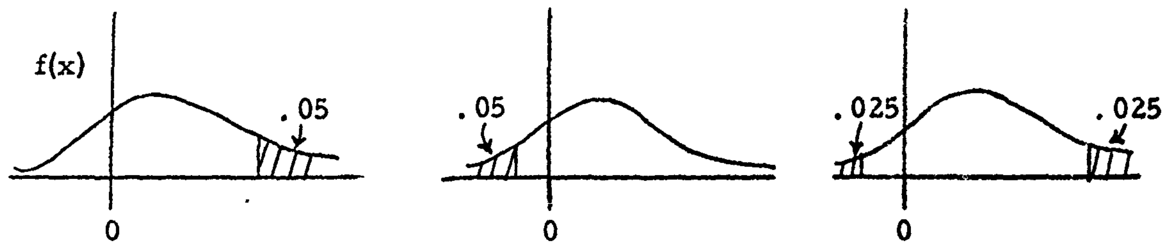


Figure 7

Some examples may now be in order.

1. Illustration. A certain stock number has the probability $p = .8$ that each requisition will be for a quantity of one unit of stock. During the past month 10,000 requisitions have been received, 8,500 were for a single unit of stock. Is the units-per-requisition pattern changing?

a. Argument. Our Null Hypothesis is that the sample behavior of the last month is consistent with the long-known $p = .8$. Now if we use the normal approximation to a binomial distribution assumed true here with $n = 10,000$ and $p = .8$, we have $\mu = 8,000$ and $\sigma = \sqrt{1,600} = 40$. Suppose we take a significance level of .01. Then we would reject the Null Hypothesis; i.e., we would reject this sample coming by chance from the assumed base population if $x > 8,000 + (2.33)(40) = 8,093$, where x = number of requisitions with a single unit per order. For this can happen in only 1% of many repeated cases due to chance. Since $8,500 > 8,093$, we reject chance and claim a significant difference due to other than chance. Hence we disassociate

sample and assumed parent population. The size-order pattern has changed. We are more than 99% confident though we do not usually so speak.

Confidence was used in reflecting how frequent due to chance something we found in one case would happen in repeated cases. When we reject, as we just did, something that was rare due to chance before it was rejected, it is not quite the same thing as making a positive confidence interval statement. Still there is a relationship.

b. Argument (Alternate). Now we could have computed a 99% confidence interval here. Since the hypothesis is $p = .8$ and, let us say, the alternative is $p > .8$, we use a one-sided confidence interval and say

$$\Pr\left\{\frac{.85 - p}{\sqrt{\frac{(.85)(.15)}{10,000}}} < 2.33\right\} = .99.$$

This can be reformed into

$$\Pr\{.835 < p\} = .99.$$

Since $p = .8 < .835$, we reject $p = .8$ as it is not in the confidence interval $(.835, 1)$.

2. Illustration. Suppose Washington decorates our unit when we are very effective and they have done this for each of the past five months. Would you say from the statistical point of view that our probability of being decorated exceeds .5?

a. Argument. From the significance testing point of view if we accept $p = .5$, then the sample situation has a probability of occurring

(75)

equal to $(.5)^5 = .031$ and we would reject the hypothesis of $p = .5$ at the 5% significance level, but not at the 1% level. Remember that if one uses a higher value for the significance level, he runs a great risk of accepting a false hypothesis.

b. Argument (Alternate). From another point of view we let x be the number of times in 5 months we are decorated and p be the probability of being decorated in any month. Then we ask that

$$\Pr\{x < 5\} < .95$$

so that

$$\Pr\{x = 5\} > .05.$$

This requires

$$p^5 > .05$$

or

$$p > .55.$$

So p must be as large as .55 to keep the sample action from being less probable than .05. Hence we do not pick up $p = .5$ in our confidence interval (.55, 1) and so we reject $p = .5$ for hypothesis. But certainly we must agree with the fact that p exceeds .5.

B. Relation between Confidence Intervals and Tests of Significance.

The practicioneer usually prefers a confidence interval statement to that using only a test of significance because the width of the confidence interval tells more about the reliance he can place on the results of the experiment. Still, when a test of significance is accompanied by the

73

appropriate Operating Characteristic Curve (OCC), about the same information is provided. In order to understand this let us first consider our situation in deciding whether to accept or to reject a hypothesis.

If we reject a hypothesis when we should not have, we say that a Type I error has been made. If we accept a hypothesis when we should not have, we say that a Type II error has been made. In general in life these two situations constitute the alternatives in making wrong decisions or errors in judgment. Ideally we want tests to minimize such errors. Unfortunately, for a fixed sample size, when we decrease one type of error we increase the other. Only increasing the sample size reduces both.

In industry in acceptance sampling the probability of the Type I error is called the Producer's Risk and denoted by α while the probability of a Type II error is called the Consumer's Risk and denoted by β. Obviously the Type I error is the basis of our familiar level of significance test. It represents the chance we are willing to take to be wrong in rejecting chance. Now we could eliminate Type II errors by never accepting hypotheses! But this would get us nowhere. Better should we study the probabilities of making Type II errors and hope for little chance of making them. The quantity $(1 - \beta)$ is helpful here in that it indicates the ability or power of the test to reject the hypothesis if it is false. Hence it is called the Power Function.

A confidence interval can be used for a test of significance--this we have illustrated. Using a rejection criterion alone in the converse

(74)

situation is not the proper way to think of a significance test. You should always think of the associated OC Curve as part of the test.

First let us look at several situations in which we highlight β.

1. Illustration. If p is the probability of a particular FSN being demanded in a day, suppose we order this item if, out of every 10 items demanded, one or more are for this FSN. The probability that the experience with 10 items does not have us order is a function of p. It is the operating characteristic function of the examination procedure and is

$$\beta = (1 - p)^{10}.$$

In Figure 8 we see a graph of β.

Figure 8

Figure 9 gives a graph of $1 - \beta$, the power function.

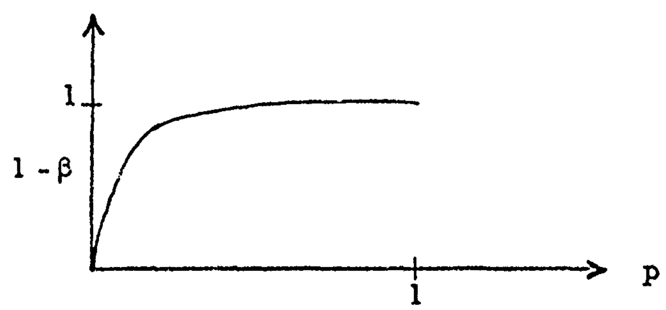


Figure 9

Now β represents the probability of not ordering the FSN when it has the probability p of being demanded and hence should be replaced according to this probability. Hence $(1 - \beta)$ represents the probability of ordering. The graphs in Figures 8 and 9 show that the procedure is pretty much in line with the actuality it is intended to follow.

2. Illustration. The Bureau of Budget (BUBUD) claims that a certain FSN is not ordered by half the Stock Points while the unit here feels it is. To test the situation you examine five Stock Points and decide to accept BUBUD's claim only if either all the five Stock Points ordered or all did not order the item. Otherwise you will assume it is ordered by half the Stock Points.

Now the probability of accepting BUBUD's claim is a function of p, the probability of a Stock Point ordering this item. This function of p is the Power Function for your test, that is, it is $(1 - \beta)$ where β is the probability of accepting your claim. Look at Figure 10.

Power $= 1 - \beta = p^5 + (1 - p)^5$

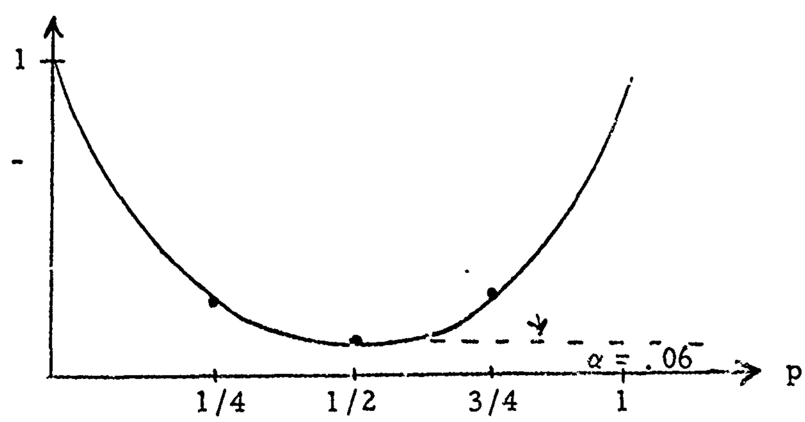


Figure 10

So if p = 1/4 or 3/4, your probability of <u>not</u> making a mistake is only about 1/4, that is, your probability of not accepting $p = 1/2$ when $p \neq 1/2$ is not very high. When $p = 1/2$, then β is the probability of accepting $p = 1/2$ when it should be accepted and so $1 - \beta$ is the probability of not accepting it when it is true. But this is the Type I error, or α, which is in this case .06.

$$[\alpha = 1 - \beta = p^5 + (1 - p)^5 = \left(\frac{1}{2}\right)^5 + \left(1 - \frac{1}{2}\right)^5 = .031 + .031 = .06]$$

This test procedure is not very powerful in that it does not strongly have you reject the hypothesis $p \neq 1/2$ when $p \neq 1/2$.

Now let us take a numerical example and tie in both approaches of confidence intervals and of significance testing with the operating characteristic curve.

3. <u>Illustration.</u> An FSN has a mean requisition size of 300 and a standard deviation of 24. Suppose we want to know at the 1% level of significance from a sample of 64 requisitions if this mean requisition size has increased.

a. <u>Argument.</u> In customary notation we say

H_0: Null Hypothesis that $\overline{D} = 300$ has not changed

H_1: alternate hypothesis that $\overline{D} > 300$ and $\overline{D}$ has changed

The 1% level of significance corresponds in normal theory to z = 2.33 and so to

$$\overline{D} = 300 + 2.33 \left(\frac{24}{\sqrt{64}}\right) = 307.0.$$

(71)

Graphically this is given in Figure 11.

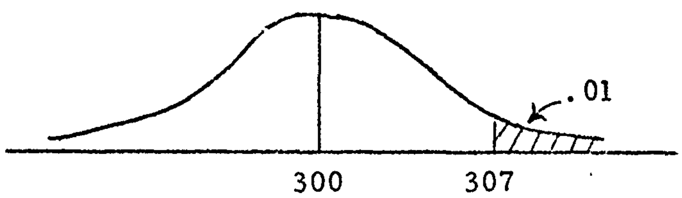


Figure 11

So our Type I error has probability $\alpha = .01$. For each possible actual new value of $\overline{D}$, there is a chance of accepting the old $\overline{D} = 300$. To show this, let us first take the value $\overline{D} = 310$ as being the actual new average demand. Then the means of samples of size 64 are normally distributed about 310 and with some chance the sample means will be to the left of the critical point point for rejecting the old hypothesis, $\overline{D} = 300$, as shown in Figure 12.

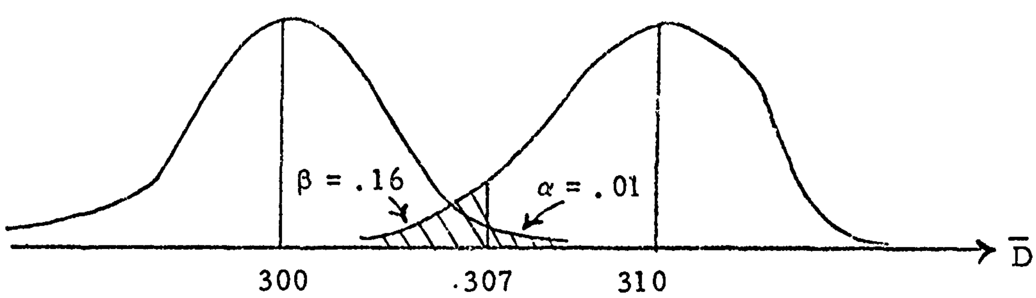


Figure 12

Now under the new hypothesis 307 corresponds to $z = -1$ and so $\beta = .16$.

More generally we can calculate β for various new $\overline{D}$ as given in Table XIV.

Table XIV

$\overline{D}$	290	295	300	305	310	315	320
β	1.00	1.00	.99	.75	.16	.00	.00

The OC Curve and Power Function are graphed in Figure 13.

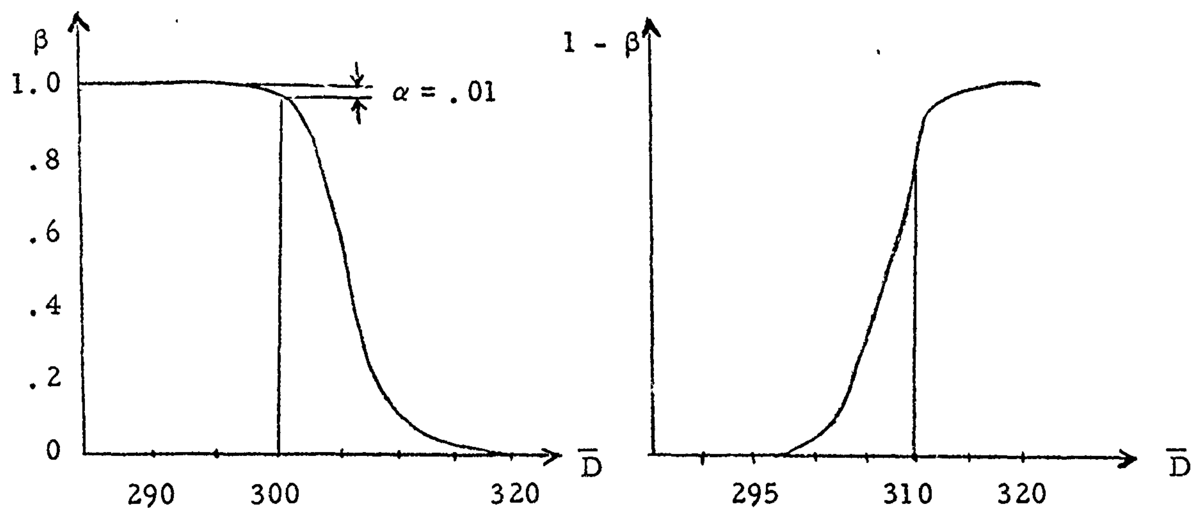


Figure 13

Next let us perform an analysis still using 99% significance similar to what we did in Illustrations 1 and 2, pages 65 and 66, and indicating the associated confidence level approach. We want and have

$$\Pr\left\{\frac{\overline{D} - \mu_{\overline{D}}}{24/\sqrt{64}} < 2.33\right\} = .99$$

which can be reformed into

71

$$Pr\{\overline{D} - 7 < \mu_{\overline{D}}\} = .99$$

or $(\overline{D} - 7, \infty)$ is the confidence interval.

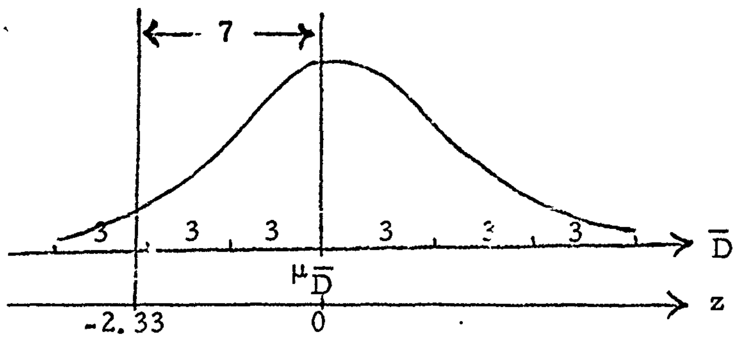


Figure 14

Note the range of 99% of the activity under the curve in Figure 14, starting 7 units to the left of $\mu_{\overline{D}}$ and thence to the right, is about 16 units which is about the span of indeterminancy in the OC Curve from 300 to 315.

We see from Figure 14 that there is a small chance of keeping the hypothesis $\overline{D} = 300$ when $\overline{D} > 315$ while we are almost certain of keeping it when $\overline{D} < 300$. Were we to have used the confidence interval approach at $c = .99$, i.e., $\alpha = .01$, our random interval would be one-sided simply because the alternate hypothesis is $\overline{D} > 300$. Thus we would say

$$(\overline{D} - 7, \infty)$$

is the confidence interval estimator and, as you know, this interval in repeated samples on the average would include $\mu_{\overline{D}} = \mu_D$ 99 out of 100 times. Also we recognize that for this same percentage of times the value of $\overline{D}$ would land in a span of about 16 units.

The two approaches can be illustrated with respect to determining the sample size in order to detect differences between means. We can specify limits to the risks for Type I and Type II errors. This locates two points on the OC Curve. Selection of n follows from examination of various OC Curves for different n and matching these two points.

On the other hand we can specify the magnitude of difference between means which is our limit. Then we can compute the sample size which gives with desired confidence an interval of this length. Let us illustrate this.

4. Illustration. In the problem of Illustration 2, page 66, suppose we test the hypothesis of the FSN being demanded half the time, i.e., $p = .5$, by examining a sample of future demands. Now let us decide

(1) the probability of rejecting $p = .5$, when it is correct, is not to exceed .05. This amounts to saying $\alpha = .05$.

(2) the probability of accepting $p = .5$ when $p \geq .6$ or $p \leq .4$ is not to exceed .05. This amounts to saying $\beta = .05$.

Find the minimum sample size and state the statistical decision rule.

a. Argument. Graphically we have Figure 15.

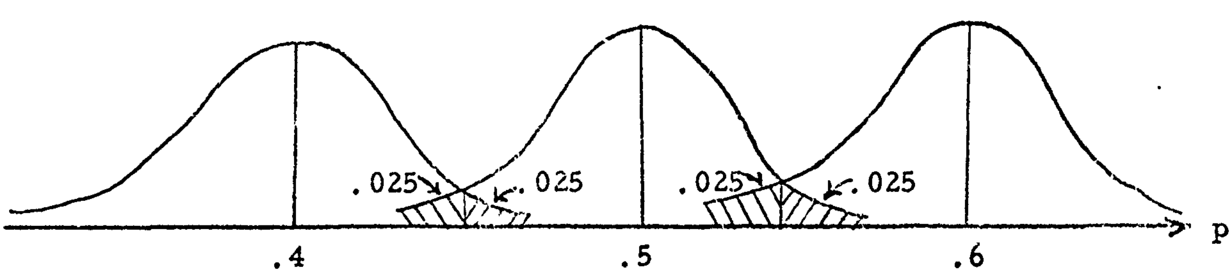


Figure 15

Let

N = number of observations in sample

x = number of times particular FSN is ordered

Then

$$\text{area under } \mu = .5 \text{ curve to right of } \frac{x - .5N}{\sqrt{N(.5)(.5)}} = .025$$

$$\text{area under } \mu = .6 \text{ curve to left of } \frac{x - .6N}{\sqrt{N(.6)(.4)}} = .025$$

$$\therefore \quad \frac{x - .5N}{.5\sqrt{N}} = 1.96 \quad \text{and} \quad \frac{x - .6N}{.49\sqrt{N}} = -1.96.$$

Hence we have the two simultaneous equations in x and N,

$$x = .5N + .980\sqrt{N}$$

$$x = .6N - .960\sqrt{N},$$

which yield

$$N = 377$$

$$x = 208.$$

When $p = .5$, then $x - Np = 19$. Thus we would get a similar span of values to the left of 189. So we decide:

(1) accept the hypothesis $p = .5$ if in 377 demands, we have demands for this FSN in the range 189 ± 19, i.e., between 170 and 208.

(2) reject the hypothesis otherwise.

5. Illustration. In the last illustration retain everything but now require: the probability of accepting $p = .5$ when actually $p \geq .6$ is .05.

In this case we find N = 319, x = 177. So we say

a. Accept p = .5 if x lies between 142 and 177.

b. Reject p = .5 otherwise.

In summary when the data present enough evidence to reject the hypothesis, the probability α of an incorrect judgment is known in advance since α is used in locating the rejection region. On the other hand, if the data present insufficient evidence to reject the hypothesis, we are not sure what to do. We should specify a practical significant alternative and calculate β. In addition if the size of a sample is involved, we should pick it so β is small. But in many practical problems the calculation of β may be difficult, if not impossible. So more often it is better not to reject rather than accept and then to estimate using a confidence interval.

V. POINT ESTIMATION

An estimate of a population parameter by a single number is called a point estimator. Historically the estimation problems concerned themselves with estimating parameters. One assumed that the distribution of probability over the base population is one of a family of distributions indexed by one or more real-valued parameters. Then estimates of the parameters were made on the basis of experimental observations.

Suppose $(x_1, x_2, \cdots, x_n)$ is a random sample from a distribution which is characterized by an unknown parameter θ. Now θ could be the mean. What we try to do in point estimation is to develop a function of the sample $(x_1, x_2, \cdots, x_n)$ which will have a distribution that will cluster about θ. More precisely a point estimator for θ is a real single-valued function of $(x_1, x_2, \cdots, x_n)$, say $t(x_1, x_2, \cdots, x_n)$ whose distribution "clusters in some sense" around θ. This t-function is itself a random variable. Graphically we like to get a distribution as shown in Figure 16.

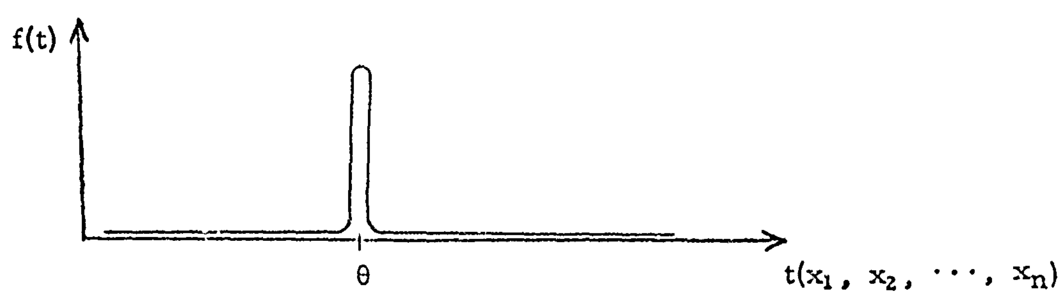


Figure 16

Our job is to try to define the phrase "in some sense," i.e., to qualify it. The present jargon used in doing this begins with two statements:

t isn't offset ⟷ unbias ,

t is as narrow as possible ⟷ efficient .

In addition there are other qualifications we will discuss, namely being consistent and sufficient.

A. Unbiasedness.

Suppose $(x_1, x_2, \cdots, x_n)$ is a random sample from a distribution $f(x)$ and suppose that there is a parameter θ which (partially) describes $f(x)$. Let $t(x_1, x_2, \cdots, x_n)$ be a random variable such that

$$E(t(x_1, x_2, \cdots, x_n)) = \theta$$

where the expectation is taken over all possible random samples. Then $t(x_1, x_2, \cdots, x_n)$ is called an unbiased estimator for θ. Precisely, the average value of t is θ.

1. Illustration. Suppose $(x_1, x_2, \cdots, x_n)$ is a random sample from a distribution $f(x)$ whose mean is θ, i.e.,

$$E(x) = \theta.$$

Let $t(x_1, x_2, \cdots, x_n) = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{x}$. Then $\bar{x}$ is an unbiased estimator for θ, as you already know, since

$$\begin{aligned} E(t) = E(\bar{x}) &= E\left(\frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n}\right) \\ &= E\left(\frac{x_1}{n}\right) + E\left(\frac{x_2}{n}\right) + \cdots + E\left(\frac{x_n}{n}\right) \\ &= \frac{1}{n}E(x_1) + \frac{1}{n}E(x_2) + \cdots + \frac{1}{n}E(x_n) \end{aligned}$$

$$= \frac{1}{n}\theta + \frac{1}{n}\theta + \cdots + \frac{1}{n}\theta$$

$$= \theta, \text{ free of } n.$$

2. Illustration. Consider a sequence of Bernoulli trials and the resulting binomial distribution of probability for the occurrence of the event of interest x times, i.e., on x number of trials. Then if $\hat{p}$ is the relative frequency sample estimate of p, we know from $E(x) = np$ that

$$E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{np}{n} = p, \text{ free of } n.$$

Therefore $\hat{p}$ is an unbiased estimate for p.

3. Illustration. Suppose we again have the situation in the previous illustration, only now we are interested in estimating the ratio $p/(1 - p)$. This ratio is often desired where the ratio of the proportions of two things is of interest. Suppose we consider samples (x_1, x_2) of size 2. Then the binomial variable which counts occurrences of the event of interest can be either 0, 1, or 2. Let $q = 1 - p$. Suppose our estimator for p/q is $t(x_1, x_2)$. Assuming it is symmetric in x_1 and x_2, we can further assume that t takes on only three different values, one for each of the three values of the binomial variable x. Call these values a, b, and c, respectively. They occur, as you know, with probabilities, q^2, $2pq$, and p^2, respectively. Then the expected value of t over all samples of size 2 is

$$\begin{aligned} E(t(x_1, x_2)) &= t(0, 0)\Pr\{t(0, 0)\} + t(0, 1)\Pr\{t(0, 1)\} \\ &\quad + t(1, 0)\Pr\{t(1, 0\} + t(1, 1)\Pr\{t(1, 1)\} \\ &= aq^2 + 2bpq + cp^2. \end{aligned}$$

But if t is an unbiased estimator for p/q, this must equal p/q, regardless of the value of p. Now

$$aq^2 + 2bpq + cp^2 \leq a + 2b + c,$$

yet we can always find a value of p close enough to 1 so that

$$a + 2b + c < p/q.$$

Hence there is no unbiased estimator for p/q in general when $n = 2$. By a similar argument it can be shown that the same conclusion is true for any other value of n.

But this simply means that we cannot find one set of numbers $\{a, b, c\}$ that works for every p. We still might be able to find a correct set when p is known. Practically this is no help, however, since our sampling problems are directed to finding p.

This last illustration is not to be regarded with too much sorrow. For though unbiasedness is a desirable property, it is not essential. An estimate that is slightly bias but very closely clustered could be more useful than an unbiased one that is widely spread. Moreover, as we shall show, if consistency exists, we know the bias disappears as the size of the sample increases. In the last illustration we know when n is large, x/n should be near p and $(n - x)/n$ near q. Hence their ratio which reduces to $x/(n - x)$ should be near p/q. In a later section this can be defended by showing it is consistent. Still you will note that this statistic defies having its expected value calculated for any fixed n. For when $n = 2$, we get

$$E\left(\frac{x}{2-x}\right) = \frac{0}{2-0}q^2 + \frac{1}{2-1}2pq + \frac{2}{2-2}p^2 = \infty.$$

We recall the median of a sample unbiasedly estimates the population median. Also we now have another good reason for the definition of standard deviation or variance of a sample, with the $n - 1$ instead of n in the denominator because

$$E\{s^2\} = E\left\{\sum_1^n \frac{(x_i - \bar{x})^2}{n-1}\right\} = \frac{1}{n-1}E\left\{\sum_1^n (x_i - \bar{x})^2\right\}$$

and since

$$\sum_1^n (x_i - \bar{x})^2 \equiv \sum_1^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

we have

$$E\left\{\sum_1^n (x_i - \bar{x})^2\right\} = E\left\{\sum_1^n (y_i - \mu)^2\right\} - E\{n(\bar{x} - \mu)^2\}$$

$$= \sum_1^n E\{(x_i - \mu)^2\} - nE\{(\bar{x} - \mu)^2\}$$

$$= n\sigma^2 - n \times \frac{\sigma^2}{n}$$

$$= (n-1)\sigma^2$$

$$\therefore \quad E\{s^2\} = \frac{1}{n-1} \times (n-1)\sigma^2 = \sigma^2, \text{ free of } n.$$

B. Efficiency.

Suppose $t(x_1, x_2, \cdots, x_n)$ and $t^*(x_1, x_2, \cdots, x_n)$ are two unbiased estimators for the parameter θ with variances σ_t^2 and $\sigma_{t^*}^2$, respectively. Then the one with the smaller variance is called an efficient estimator

for θ while the other is called an inefficient one. This is rather loose talk and we should use modifiers of a comparative nature. Equivalently, but in another way, we say the efficiency of t relative to t^* for estimating θ is

$$\sigma^2_{t^*}/\sigma^2_t .$$

So when efficiency is less than 1, the other statistic is more efficient. When a value of an efficient statistic is given, it is called an efficient estimate.

The following theorem is remarkable and has been known a long time.

1. Theorem. Let $(x_1, x_2, \cdots, x_n)$ be a random sample from a distribution whose mean is θ. Consider the weighted mean

$$\bar{x}_w = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

where $c_1 + c_2 + \cdots + c_n = 1$. Then $\bar{x}_w$ is an unbiased estimator for θ and the variance of it attains its minimum value when $c_1 = c_2 = \cdots = c_n = 1/n$.

a. Argument. Consider $n = 2$. Then (x_1, x_2) is our random sample from a population whose mean is θ and variance is σ^2. Now

$$\bar{x}_w = c_1 x_1 + c_2 x_2 , \qquad c_1 + c_2 = 1$$

$$\sigma^2_{\bar{x}_w} = c_1^2\sigma^2 + c_2^2\sigma^2 = (c_1^2 + c_2^2)\sigma^2 = k^2 < \sigma^2 .$$

Geometrically we are considering only points on the line $c_1 + c_2 = 1$ and also on the circle $c_1^2 + c_2^2 = (k/\sigma)^2$ as seen in Figure 17.

In order to get the smallest value of k/σ and still get a point of intersection with the line $c_1 + c_2 = 1$, we want the circle to just touch the

line. Then the radius is $1/\sqrt{2}$. Thus k/σ is $1/\sqrt{2}$. At this point $c_1 = c_2 = 1/2$ which says $\bar{x}_w$ should have the particular value $\bar{x}$ for minimum variance which then is $\sigma^2/2$.

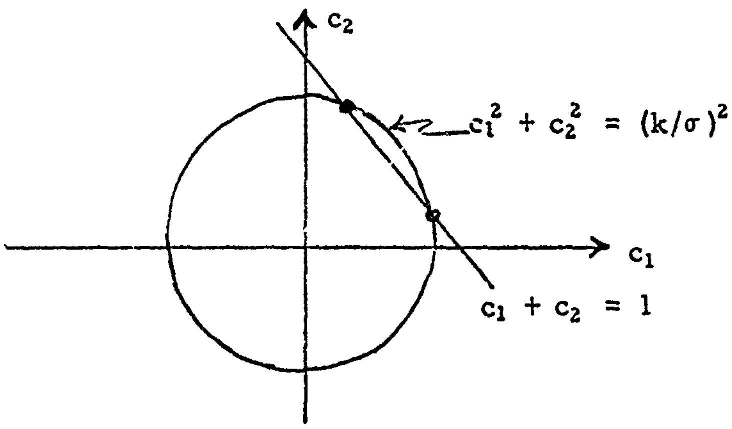


Figure 17

This means if you take a sample, you can't expect to do any better than to take its mean to estimate the mean of the base population so far as variance is concerned.

Thus the mean of the sample is called the <u>most efficient estimator</u> for the base mean.

The forecast of quarterly demand is a point estimator. Locally it is developed using single exponential smoothing and past observations of demand. The demands are weighted, but the weights decrease geometrically. The reason for this technique is to give greater emphasis to most recently experienced demand. The sum of the weights applied to demand observations does not equal one because the last term in the formula for single exponential smoothing contains a previous forecast and it is also weighted.

The sum of all weighting factors does equal one, however. The equation for single exponential smoothing can be written as follows:

$$\bar{x}_0 = c_1 x_1 + c_2 x_2 + \cdots + c_m x_m + (1 - c_1)^m \bar{x}_m$$

where

$$c_2 = c_1(1 - c_1)$$

$$c_3 = c_1(1 - c_1)^2$$

and

$$c_m = c_1(1 - c_1)^{m-1}$$

or

$$c_i = c_1(1 - c_1)^{i-1}$$

and c_1 is always a positive fraction.

The variance of $\bar{x}_0$ is determined as follows:

$$\bar{x}_0 = c_1 x_1 + c_1(1 - c_1)x_2 + \cdots + c_1(1 - c_1)^{m-1} x_m + (1 - c_1)^m \bar{x}_m$$

$$\begin{aligned}
\sigma^2_{x_0} &= c_1^2 \sigma^2_{x_1} + c_1^2(1 - c_1)^2 \sigma^2_{x_2} + \cdots + c_1^2(1 - c_1)^{2(m-1)} \sigma^2_{x_m} \\
&\quad + (1 - c_1)^{2m} \sigma^2_{\bar{x}_m} \\
&= [c_1^2 + c_1^2(1 - c_1)^2 + \cdots + c_1^2(1 - c_1)^{2(m-1)}] \sigma^2_x \\
&\quad + (1 - c_1)^{2m} \sigma^2_{\bar{x}_m} \\
&= c_1^2 \left[\frac{1 - (1 - c_1)^{2m}}{1 - (1 - c_1)^2} \right] \sigma^2_x + (1 - c_1)^{2m} \sigma^2_{\bar{x}_m} \\
&= \frac{c_1}{2 - c_1} [1 - (1 - c_1)^{2m}] \sigma^2_x + (1 - c_1)^{2m} \sigma^2_{\bar{x}_m}
\end{aligned}$$

$$\rightarrow \frac{c_1}{2 - c_1} \sigma_x^2 \text{ as } m \rightarrow \infty$$

but

$$\sigma^2_{n\text{-period average}} = \frac{\sigma_x^2}{n}$$

so single exponential smoothing is as efficient as an n-period average where $n = (2 - c_1)/c_1$. Graphically:

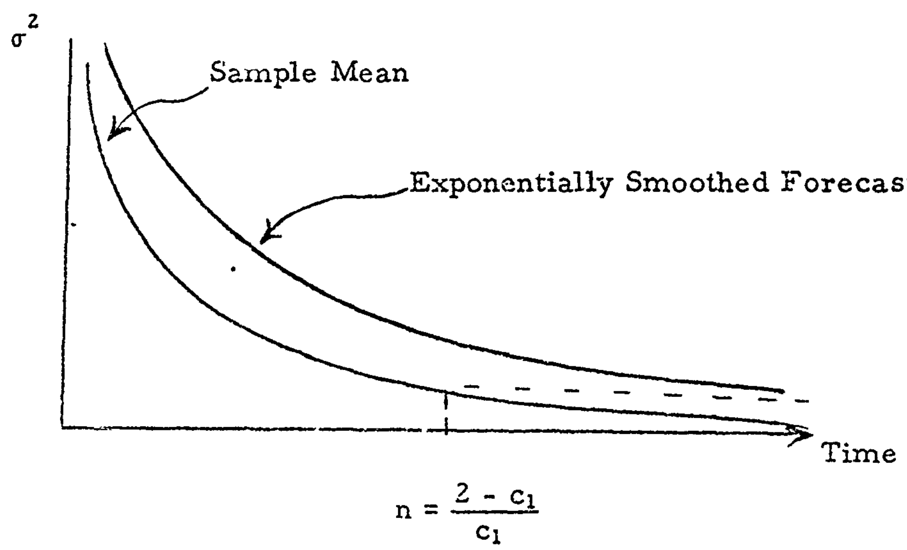


Figure 18

Depending on the efficiency of $\bar{x}_m$, the forecast developed by the smoothing technique can theoretically be no better than an n-period mean. Of course, this conclusion assumes the demand distribution does not change.

2. Illustration 1. Let $(x_1, x_2, \cdots, x_n)$ be a random sample from $N(\mu, \sigma^2)$ and let $\bar{x}$ be the mean and $\tilde{x}$ be the median. You know that $\sigma_{\bar{x}}^2 = \sigma_x^2/n$. Now we can show that

$$\sigma_{\tilde{x}}^2 = \frac{\pi}{2}\frac{\sigma_x^2}{n} + O\left(\frac{1}{n^{3/2}}\right) .$$

Since $\bar{x}$ is the best estimator for μ in any distribution, the efficiency of $\tilde{x}$ relative to $\bar{x}$ for estimating μ is

$$\frac{\sigma^2/n}{\frac{\pi}{2}\sigma^2/n} = \frac{2}{\pi} = .64.$$

This means a sample of size 64 is just as good when taking the arithmetic mean as is one of size 100 when using the median.

3. Illustration 2. For a unit standard normal distribution, N(0, 1), we find the average deviation from the mean is $\sqrt{2/\pi} = 0.79788 \doteq .8$. Further we find in sampling from such a distribution, the Mean Absolute Deviation (MAD) of the sample is an unbiased estimator for the base population MAD. Hence we have

$$E(1.25\text{MAD}) = \sigma$$

which explains our correction to the "PROGRAM 61" calculation of MAD to estimate σ.

How efficient then is the MAD? Well, the variance of 1.25MAD from samples of size n is

$$\sigma^2_{(1.25\text{MAD})} = \text{var}\left[\sqrt{\frac{\pi}{2}}\sum\frac{|x_i - \mu|}{n}\right]$$

$$= \frac{\pi}{2n^2}\Sigma\,\text{var}\,|x_i - \mu|$$

$$= \frac{\pi}{2n^2} \times n \times \text{var}\,|x - \mu|$$

$$= \frac{\pi}{2n}\{E|x - \mu|^2 - [E|x - \mu|]^2\}$$

$$\doteq \frac{\pi}{2n}\{\sigma^2 - \sigma^2 \times \frac{2}{\pi}\} = \frac{\pi - 2}{2n}\sigma^2 .$$

whereas the variance of the standard deviation s' from samples of size n (corrected from s so that $E(s') = \sigma$) is

$$\sigma^2_{s'} = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right)$$

Therefore the efficiency of 1.25MAD relative to s' is

$$\text{Eff} = \frac{\frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right)}{\frac{\pi - 2}{2n}\sigma^2} \cong \frac{1}{\pi - 2} = .8760.$$

In practice one usually does not have μ and resorts to using $\bar{x}$ in its place. In such cases our formula for σ in terms of the MAD must be corrected to

$$\sqrt{\frac{\pi}{2}}\sum \frac{|x_i - \bar{x}|}{\sqrt{n}\,\sqrt{n - 1}}$$

so that it is unbiased. A similar correction is needed for s. Remember we only compare for relative efficiency the variance of two statistics, when each is an unbiased estimator of the same parameter. Going back to the correction for the sample MAD estimate of σ, we see from

$$\frac{1}{\sqrt{n(n - 1)}} = \frac{1}{n} + \frac{1}{2n^2} + \frac{3}{8n^3} + \frac{15}{48n^4} + \cdots$$

that

$$\Delta = \frac{1}{\sqrt{n(n - 1)}} - \frac{1}{n}$$

gives .0054 for $n = 10$ and .0008 for $n = 25$. Hence for all practical purposes this correction is never of importance.

Fisher remarked on these two estimators: "As n is made larger, therefore the standard error of 1.25MAD tends to bear a constant ratio to that of s. The former is the larger in the ratio $\sqrt{\pi - 2}$; in other words, the value of the standard deviation obtained from s^2 of a sample has greater weight by 14% than that obtained from 1.25MAD. To obtain a result of equal accuracy by the latter method, the number of observations must be increased by 14%."

C. Consistency.

We say $t(x_1, x_2, \cdots, x_n)$ is a consistent estimator for θ if

$$\lim_{n\to\infty} \Pr\{|t(x_1, x_2, \cdots, x_n) - \theta| < \delta\} = 1 \quad \text{for any } \delta > 0.$$

Fisher called this the common-sense criterion and stated it as follows: When applied to the whole population the derived statistic should be equal to the parameter. This means as n gets bigger all the probability of the distribution of the statistic t lies in the interval $(\theta - \delta, \theta + \delta)$. This convergence in probability to a constant is also convergence in distribution. From the graph in Figure 19 we see we could equivalently say:

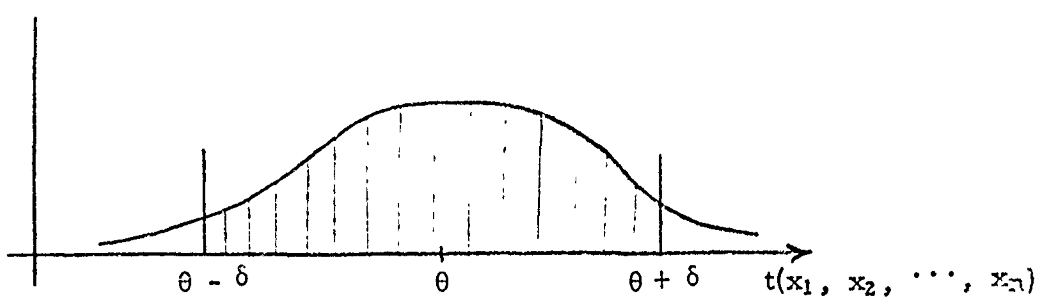


Figure 19

for an arbitrary $\delta > 0$ and $\epsilon > 0$, no matter how small, we can find an $n(\delta, \epsilon)$ such that

$$\Pr\{\theta - \delta < t(x_1, x_2, \cdots, x_n) < \theta + \delta\} > 1 - \epsilon \quad \text{for } n > n(\delta, \epsilon).$$

1. Theorem. Let $(x_1, x_2, \cdots, x_n)$ be a sample from a distribution whose mean is θ and variance is σ^2. Then $\bar{x}$ is a consistent estimator for θ.

a. Argument. We know $E(\bar{x}) = \theta$ and $\sigma_{\bar{x}}^2 = \sigma_x^2/n$. In Tchebycheff's inequality

$$\Pr\{|\bar{x} - \theta| < \lambda\sigma_{\bar{x}}\} > 1 - 1/\lambda^2$$

let $\delta = \lambda\sigma_{\bar{x}}$. Then we obtain

$$\Pr\{|\bar{x} - \theta| < \delta\} > 1 - 1/(\delta/\sigma_{\bar{x}})^2 = 1 - \sigma_x^2/n\delta^2.$$

Hence

$$\lim_{n\to\infty} \Pr\{|\bar{x} - \theta| < \delta\} = 1.$$

You can see the property of consistency is concerned with the behavior of an estimator when the number n of elements in the outcome is large. Actually we have used the Law of Large Numbers on several occasions to show that an estimate is consistent, for example, $\bar{x}$ for μ and s_x^2 for σ_x^2. Then again we showed this for proportions with respect to probabilities in the case of the binomial distribution. However it is possible in this case to get a strong conviction for it by a more detailed examination such as that given in Appendix A.

In general, if $t(x_1, x_2, \cdots, x_n)$ is an unbiased estimator for θ and $\sigma_t^2 \to 0$ as $n \to \infty$, we know the estimates more closely approach θ as n increases.

D. Sufficiency.

In 1920 Fisher became impressed by what he called the characteristic of sufficiency. He assumed a normal base distribution with standard deviation σ. Then he considered the two common methods of estimating σ^2 or σ from a sample $(x_1, x_2, \cdots, x_n)$, namely

$$n\sigma_1 = \sqrt{\frac{\pi}{2}}\, S(|x - \bar{x}|) \qquad \text{Mean Error}$$

$$n\sigma_2^2 = S(x - \bar{x})^2 \qquad \text{Mean Square Error,}$$

sometimes called Peter's formula and Bessel's formula, respectively. He showed the ratio of the variances for σ_1 and for σ_2 to be $(\pi - 2)$, as we discussed in a preceding section. Then he considered various powers p of the deviations and showed the precision of the mean square is a true maximum, i.e., for $p = 2$, while the variance is 14% greater for $p = 1$ and 9% for $p = 3$. Hence we have still another good reason for preferring Bessel's formula.

But even more important he showed that for a given value of σ_2 the distribution of σ_1 is independent of σ. So when σ_2 is known, a value of σ_1 can give no additional information as to the true value of σ. The same can be said if any other estimator is substituted for σ_1. Consequently the whole of the information concerning the base population variance which a sample provides is summed up in the single estimate σ_2. Now the same cannot be said for σ_1 being taken first, since then σ_2 does involve σ. This means we could improve our estimate of σ when we first determine σ_1 by taking σ_2.

92

One must remember that this unique superiority of σ_2 depends on the normal curve hypothesis for the base distribution. For some other curve, σ_1 might be the superior estimator for σ. As a matter of fact it is when the base population is of the form

$$\frac{1}{\sigma\sqrt{2}}\, e^{-\frac{|x-m|\sqrt{2}}{\sigma}} ,$$

a double exponential curve. In this case σ_1 must be altered to

$$n\sigma_1 = \sqrt{2}\, S(|x - \bar{x}|).$$

Fisher suggested we calculate β_2, the ratio of the fourth moment to the square of the second moment. If this is near 3, thé Mean Square Error should be used; if this is near 6, perhaps we would be better using σ_1 for our estimate of σ.

Later we will see that when this property of sufficiency exists for an estimator, we will be able in general to find the estimator by the Method of Maximum Likelihood. Also such a statistic will be most efficient if a most efficient estimator exists.

The usual academic form in which the criterion of sufficiency is presented leaves a lot to be desired insofar as determining a sufficient estimator. The ordinary definition requires you know the statistic before its sufficiency can be tested. This is why Fisher said he provided us with the Method of Maximum Likelihood--to provide a statistic for which the criterion of sufficiency is satisfied.

To exemplify this concept we shall examine several situations.

1. Illustration. Consider the mean of the Poisson distribution

$$\frac{e^{-m} m^x}{x!}$$

The parameter m may be estimated from the mean $\bar{x}$ of the observed sample. Now it can be proved that the distribution of $n\bar{x}$ is again the Poisson series

$$\frac{e^{-nm}(nm)^{n\bar{x}}}{(n\bar{x})!}.$$

The probability of drawing in order any particular sample $(x_1, x_2, \cdots, x_n)$ is

$$e^{-nm} \frac{m^{n\bar{x}}}{x_1! x_2! \cdots x_n!}$$

and this may be divided into two factors, viz.,

$$e^{-nm} \frac{(nm)^{n\bar{x}}}{(n\bar{x})!} \times \frac{(n\bar{x})!}{x_1! x_2! \cdots x_n!} \left(\frac{1}{n}\right)^{x_1} \left(\frac{1}{n}\right)^{x_2} \cdots \left(\frac{1}{n}\right)^{x_n}$$

of which the first factor represents the probability that the actual total $n\bar{x}$ should have been scored, and the second factor the probability, given this total, that the partition of it among the n observations should be that actually observed. In the latter factor, m, the parameter sought, does not appear. Hence $\bar{x}$ is a sufficient statistic for m.

a. Definition. Suppose a population has a probability density $f(x, \theta)$, where θ is a parameter. Let $(x_1, x_2, \cdots, x_n)$ be a random sample. If $t(x_1, x_2, \cdots, x_n)$ is a function (random variable with its own probability law) such that the probability density function of $(x_1, x_2, \cdots, x_n)$

for any fixed value of $t(x_1, x_2, \cdots, x_n)$ does not depend on θ, then $t(x_1, x_2, \cdots, x_n)$ is a sufficient statistic and it or some simple function of it will be a sufficient estimator for θ.

This means that if

$$g((x_1, x_2, \cdots, x_n) \mid t(x_1, x_2, \cdots, x_n)) = t'$$

does not depend on θ, then t is sufficient.

2. Illustration. Let

$$f(x; \theta) = \begin{cases} \theta e^{-x\theta} & , x \geq 0 \\ 0 & , x < 0. \end{cases}$$

Take a random sample $(x_1, x_2, \cdots, x_n)$. It has probability density function

$$\theta^n e^{-\theta \Sigma x_i}$$

Let $t(x_1, x_2, \cdots, x_n) = \Sigma x_i$. Then

$$\theta^n e^{-\theta t} = \theta \frac{(\theta t)^{n-1} e^{-\theta t}}{(n-1)!} \times \frac{(n-1)!}{t^{n-1}}$$

$$= [p(t, \theta)] \times [g((x_1, x_2, \cdots, x_n) \mid t)].$$

Since $g(x_1, x_2, \cdots, x_n \mid t)$ is the constant $(n-1)!/t^{n-1}$, we know t is sufficient. In this case we can see the geometry for small size samples, viz., as in Figure 20.

This idea of sufficiency essentially says that in the space of samples $(x_1, x_2, \cdots, x_n)$ you take a "slice" in such a way that there is a fixed probability all over this slice. Then any function of sample values on this slice has nothing to do with the parameter. All information about the parameter is obtained by going from one slice to another and not within a

slice. Incidentally, unbiasedness is not related to this. In the last illustration as well as in many others we could take Σx_i or $\Sigma x_i/n$ for a sufficient statistic. Usually a simple transformation makes it a sufficient estimator. Unfortunately, sufficient estimators are the exceptions rather than the rule. In practice we have to be content with less satisfactory estimators. However when a sufficient estimator exists and a most efficient estimator exists, we know the sufficient estimator is most efficient.

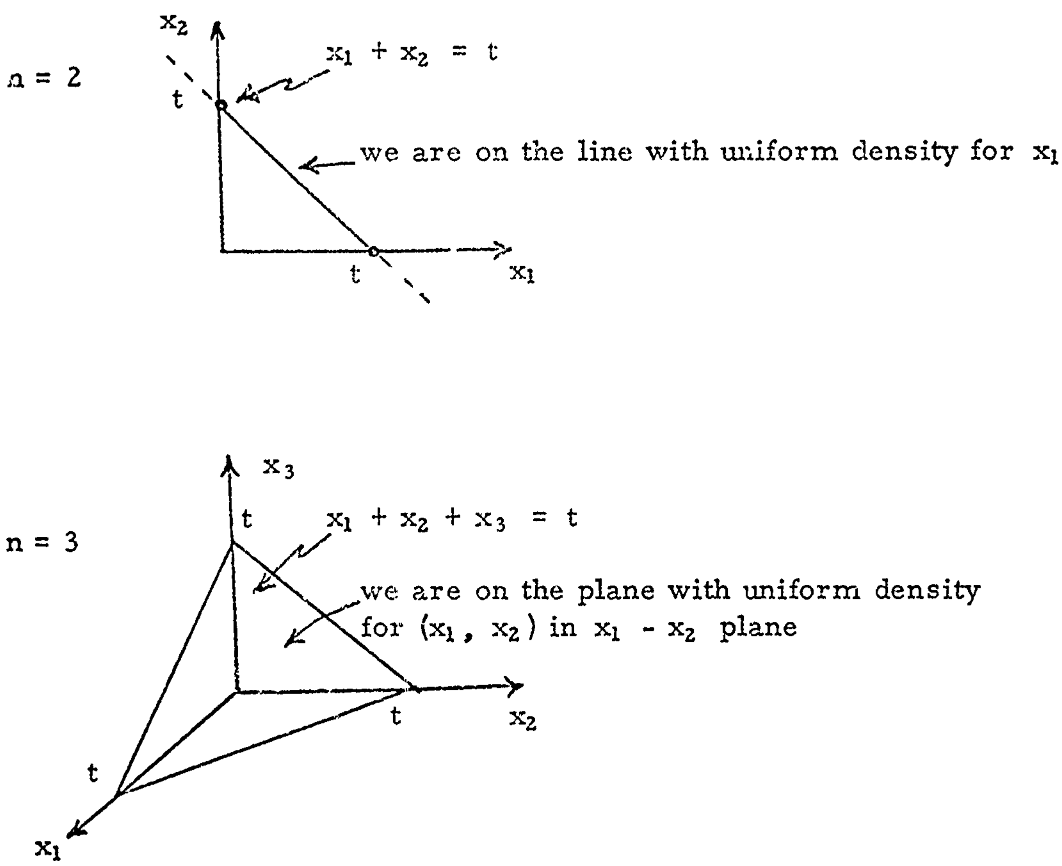


Figure 20

3. <u>Illustration.</u> Let $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$. Then for a random sample $(x_1, x_2, \cdots, x_n)$ we have

$$f(x_1, x_2, \cdots, x_n) = f(x_1, \theta) \times f(x_2, \theta) \cdots f(x_n, \theta)$$

$$= \theta^{\Sigma x_i}(1 - \theta)^{n - \Sigma x_i}$$

$$= \left[\binom{n}{t}\theta^t(1 - \theta)^{n-t}\right] \times \left[\frac{1}{\binom{n}{t}}\right]$$

$$= [p(t, \theta)] \times [g((x_1, x_2, \cdots, x_n) | \Sigma x_i = t)]$$

which again shows $t = \Sigma x_i$ is sufficient for θ.

So we see that if a statistic t is sufficient for θ, it means that the conditional distribution of any other statistic y, given $t = t'$, does not depend on the parameter θ. Consequently when we know $t = t'$, it is impossible to use y to make a statistical inference about θ. For example, you cannot then use y to find a confidence interval for θ. We might try to show that $\bar{x}$ is a sufficient estimator for the mean μ of the distribution $N(\theta, 1)$.

4. <u>Illustration</u>. Let

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}.$$

Then

$$f(x_1, x_2, \cdots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\Sigma(x_i - \theta)^2/2}.$$

Now if we expand the numerator in the exponent on e, we get

$$\Sigma x_i^2 - 2n\bar{x}\theta + n\theta^2.$$

Next if we use the identity $\Sigma x_i^2 = n\bar{x}^2 + \Sigma(x_i - \bar{x})^2$ to replace Σx_i^2 in the last expression, it becomes

$$n\bar{x}^2 + \Sigma(x_i - \bar{x})^2 - 2n\bar{x}\theta + n\theta^2$$
$$= \Sigma(x_i - \bar{x})^2 + n\bar{x}^2 - 2n\bar{x}\theta + n\theta^2$$
$$= n(\bar{x} - \theta)^2 + \Sigma(x_i - \bar{x})^2.$$

Therefore

$$f(x_1, x_2, \cdots, x_n; \theta) = \left[\frac{1}{\sqrt{2\pi}\,\frac{1}{\sqrt{n}}} e^{-\frac{(\bar{x}-\theta)^2}{2/n}}\right]$$
$$\times \left[\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \frac{1}{\sqrt{n}} e^{-\frac{\Sigma(x_i-\bar{x})^2}{2}}\right]$$
$$= [p(\bar{x}; \theta)] \times \left[g\left((x_1, x_2, \cdots, x_n) \,\middle|\, \bar{x} = \frac{\Sigma x_i}{n}\right)\right]$$

and so $\bar{x}$ is a sufficient estimator for θ.

E. Maximum Likelihood.

In 1922 Fisher introduced his Method of Maximum Likelihood to provide a statistic that was sufficient. The likelihood function L is the compound probability function or density function in the case of a continuous distribution of a specific observed sample, i.e.,

$$L = f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta)$$

for a sample $(x_1, x_2, \cdots, x_n)$ from $f(x; \theta)$. Since the logarithm of L is maximum for the same value of θ that maximizes L and since the logarithm of a product is easier to differentiate, we set

$$\frac{\partial}{\partial\theta} \log L$$

equal to zero and solve for θ in terms of our sample values. Note this L is not a probability as it does not obey the laws of probability with respect to θ.

When $\frac{\partial}{\partial\theta} \log L$ is the same function for all samples yielding the same estimate $\hat{\theta}$, then a sufficient statistic exists.

The condition that $\partial L/\partial\theta$ should be constant over the same sets of samples for all values of θ, which has been shown to establish the existence of a sufficient estimate of θ, thus requires that the likelihood is a function of θ, which, apart from a factor dependent on the sample, is of the same form for all samples yielding the same estimate $\hat{\theta}$. The sufficiency of sufficient statistics may thus be traced to the fact that in such cases the value of $\hat{\theta}$ itself alone determines the form of the likelihood as a function of θ.

1. Illustration. A sample $(x_1, x_2, \cdots, x_n)$ of n demands come at random from the exponential distribution

$$f(x) = ke^{-kx}, \qquad 0 < x < \infty.$$

Then

$$L = k^n e^{-k(x_1+x_2+\cdots+x_n)},$$

$$\frac{\partial}{\partial k} \ln L = \frac{n}{k} - (x_1 + x_2 + \cdots + x_n)$$

which when set equal to zero gives

$$\hat{k} = n/(x_1 + x_2 + \cdots + x_n) = 1/\bar{x}.$$

Thus the sample mean $\bar{x}$ is the maximum likelihood estimator for the population mean $1/k$.

2. Illustration. Suppose the random sample $(x_1, x_2, \cdots, x_n)$ comes from the normal distribution $N(\mu, \sigma^2)$. Further suppose

a. σ is known and μ is unknown. Then

$$L = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left[-\sum \frac{(x_i - \mu)^2}{2\sigma^2}\right],$$

$$\ln L = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \Sigma(x_i - \mu)^2,$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \Sigma(x_i - \mu)$$

which when set equal to zero yields

$$\hat{\mu} = (\Sigma x_i)/n = \overline{x}.$$

b. μ is known and σ is unknown. Then

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\Sigma(x_i - \mu)^2}{\sigma^3}$$

which when set equal to zero yields

$$(\hat{\sigma})^2 = \Sigma(x_i - \mu)^2/n.$$

c. μ and σ are both unknown. Now we must solve simultaneously

$$\begin{cases} \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4} \Sigma(x_i - \mu)^2 = 0 \\[2ex] \dfrac{\partial \ln L}{\partial \mu} = -\dfrac{1}{2\sigma^2} \Sigma\, 2(x_i - \mu)(-1) = 0 \end{cases}$$

From the second equation we get

$$\hat{\mu} = (\Sigma x_i)/n = \overline{x}.$$

Substituting this $\overline{x}$ for μ in the first equation yields

$$\widehat{\sigma^2} = \Sigma(x_i - \overline{x})^2/n.$$

Note the estimate $\hat{\mu}$ is unbiased but the estimate $\widehat{\sigma^2}$ is biased. However by multiplying by the constant factor $n/(n - 1)$ we can make the latter estimate unbiased. Incidentally the corresponding estimate for $\hat{\sigma}$ in 2b is not biased since μ and not $\overline{x}$ is subtracted from each x_i.

F. Relation of Maximum Likelihood to Sufficiency.

For unbiased estimators you need consider only those estimates based on (but necessarily equal to) a sufficient statistic. The sufficient statistic may be a biased estimate, but this is easily adjusted as you have seen. The remarkable thing is that for many problems there is at most one unbiased estimator based on a sufficient statistic.

Now if a problem has a sufficient statistic, then the maximum likelihood estimator is based on that sufficient statistic. Before showing this, let us recognize an alternate definition of sufficiency in the

1. Theorem. A statistic $t(x_1, x_2, \cdots, x_n)$ is sufficient for the one-parameter family $f(x; \theta)$ if and only if the sample probability function or probability density function can be factored

$$f(x_1, x_2, \cdots, x_n; \theta) = p(t; \theta) \times k(x_1, x_2, \cdots, x_n)$$

into two parts (two distributions often), one dependent only on the statistic and the parameter, the second independent of the parameter.

We can state this more generally for two or more parameters. Though we have already said this "hunt-and-peck" system is not desirable for locating a sufficient statistic, it is for the moment to be recognized that the factorization says that the variation of the probability with θ is

tied to the statistic t, and that any other variation is independent of θ. Now let's use this to show the

2. Theorem. If a sufficient statistic exists, then the maximum likelihood estimate is based on it.

a. Argument. Let $t(x_1, x_2, \cdots, x_n)$ be the sufficient statistic. Then we know by hypothesis that

$$f(x_1, x_2, \cdots, x_n; \theta) = p(t; \theta)\, h(x_1, x_2, \cdots, x_n).$$

The equation for the maximum likelihood estimate is

$$\frac{\partial}{\partial\hat{\theta}}[p(t; \hat{\theta})\, h(x_1, x_2, \cdots, x_n)] = 0$$

or

$$\frac{\partial}{\partial\hat{\theta}}\, p(t; \hat{\theta}) = 0$$

which, when solved for the maximizing $\hat{\theta}$, produces an estimate that depends only on t.

C. Normality of M. L. E. for Large Samples.

Before establishing the type of the distribution of the M. L. E. (Maximum Likelihood Estimator), let us calculate two expectations.

Consider $\frac{\partial}{\partial\theta} \ln f(x; \theta)$. Note its mean value is zero, viz.,

$$E\left(\frac{\partial}{\partial\theta} \ln f(x; \theta)\right) = \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} \ln f(x; \theta) \times f(x; \theta)dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{f(x; \theta)} \times \frac{\partial f(x; \theta)}{\partial\theta} \times f(x; \theta)dx$$

$$= \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(x;\ \theta)\,dx$$

$$= \frac{\partial}{\partial\theta} \int_{-\infty}^{+\infty} f(x;\ \theta)\,dx$$

$$= \frac{\partial}{\partial\theta}(1) = 0.$$

Next consider its variance. To avoid a lot of symbols, let S stand for $\frac{\partial}{\partial\theta} \ln f(x;\ \theta)$. Then

$$\sigma_S^2 = E(S^2) = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial\theta} \ln f(x;\ \theta)\right]^2 f(x;\ \theta)\,dx.$$

This function S, its mean, and its variance play an important role in our work as we shall see with the variance of it in the next section. Right now we further realize that the sum $\Sigma S(x_i;\ \theta)$, which we set equal to zero to get the M. L. E. $\hat{\theta}$, is a sum of independent and identically distributed random variables and hence has a limiting normal distribution with 0 for its mean and $n\sigma_S^2$ for its variance. So, for large values of n, $\hat{\theta}$ is close to θ and there is an approximately linear relation between $\Sigma S(x_i;\ \theta)$ and $\hat{\theta} - \theta$, in general.

Another way of saying this is

$$E(\Sigma S(x_i;\theta)) = 0 \quad \text{and} \quad \hat{\theta} - \theta \doteq C[\Sigma S(x_i;\ \theta)]$$

$$(\Sigma S) : N(0,\ n\sigma_S^2) \quad \text{and} \quad \hat{\theta} : N\left(\theta,\ \frac{1}{n\sigma_S^2}\right)$$

Later we shall see that $\hat{\theta}$ has minimum variance. Before that, let us consider a statement of great content.

H. Information (Frechet, Cramer-Rao) Inequality.

Suppose x_1 is a sample of size one from a probability density function $f(x;\theta)$. Let $r(x_1)$ be any unbiased estimator of θ. If

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} f(x;\theta)dx = \int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta} f(x;\theta)dx\,,$$

then

$$\sigma_r^2 \geqq \frac{1}{E\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2}\,.$$

To derive this result, let us drop the subscript on x_1 and proceed as follows. By definition

$$\int[r(x) - \theta]f(x;\theta)dx = 0.$$

Differentiation of the last expression gives

$$\int[r(x) - \theta]\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)f(x;\theta)dx + \int(-1)\,f(x;\theta)dx = 0.$$

Therefore

$$\int\left[\{r(x) - \theta\}\sqrt{f(x;\theta)}\right]\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\sqrt{f(x;\theta)}\right]dx = 1.$$

Using Schwarz's Inequality which in general is

$$\int g^2(x)dx\cdot\int h^2(x)dx \geqq \left[\int g(x)\,h(x)dx\right]^2,$$

we get

$$\int[r(x) - \theta]^2 f(x;\theta)dx \times \int\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)dx \geqq 1$$

$$\therefore \quad \sigma_r^2 \geqq \frac{1}{E\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2}\,.$$

The equality holds only when $g(x)$ and $h(x)$ are linearly related. In our case this means

$$r(x) - \theta = C \frac{\partial \ln f(x;\ \theta)}{\partial \theta}, \text{ identically in } x.$$

1. <u>Theorem.</u> For a sample of size n, the last theorem extends to

$$\sigma_r^2 \geqq \frac{1}{nE\left[\frac{\partial \ln f(x;\ \theta)}{\partial \theta}\right]^2}\ .$$

and equality again means

$$C\sum_{i=1}^{n} \frac{\partial \ln f(x_i;\ \theta)}{\partial \theta} = r(x_1, x_2, \cdots, x_n) - \theta.$$

So the last two theorems tell us what the lower bound of variation is and when we can achieve it. Let us compute this in the case of the normal

$$f(x;\ \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}, \ -\infty < x < \infty.$$

Now

$$\ln f(x;\ \theta) = -\ln\sqrt{2\pi} - (1/2)(x-\theta)^2,$$

$$\frac{\partial \ln f(x;\ \theta)}{\partial \theta} = x - \theta$$

and so

$$\sum_{i=1}^{n} \frac{\partial \ln f(x_i;\ \theta)}{\partial \theta} = \sum_{i=1}^{n} x_i - n\theta$$

$$= n\left[\frac{\Sigma x_i}{n} - \theta\right]$$

$$= \text{constant}\left[\begin{matrix}\text{unbiased estimator}\\ \text{depending only on } x'\text{s}\end{matrix} - \theta\right].$$

111

Therefore we know $\bar{x}$ is the best estimator of θ in the sense of having minimum variance.

Let's look at a discrete case

$$f(x;\,\theta) = \theta^{x}(1-\theta)^{1-x}, \qquad x = 0,\ 1.$$

Then

$$\ell n\, f(x;\,\theta) = x\,\ell n\,\theta + (1-x)\,\ell n\,(1-\theta),$$

$$\frac{\partial}{\partial\theta}\,\ell n\, f(x;\,\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)},$$

and so

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\,\ell n\, f(x_i;\,\theta) = \frac{1}{\theta(1-\theta)}\left[\Sigma x_i - \theta n\right] = \frac{n}{\theta(1-\theta)}\left[\frac{\Sigma x_i}{n} - \theta\right].$$

Once again $(\Sigma x_i)/n$ shows up to be a good estimator. Since Σx_i is the number of ones or occurrences, this proportion is the best estimator for θ.

Incidentally if we know we have the lower bound of variance for our statistic, this theory can give us a quick way to calculate it. For in the normal example

$$\sigma_{\bar{x}}^{2} = \frac{1}{nE\left[\frac{\partial\,\ell n\, f}{\partial\theta}\right]^{2}} = \frac{1}{n(1)} = \frac{1}{n}$$

and in the second example

$$\sigma_{\bar{x}}^{2} = \frac{1}{nE\left[\frac{\partial\,\ell n\, f}{\partial\theta}\right]^{2}} = \frac{1}{nE\left[\frac{x}{\theta} - \frac{1-x}{1-\theta}\right]^{2}}$$

$$= \frac{1}{n\left[\frac{1}{\theta(1-\theta)}\right]} = \frac{\theta(1-\theta)}{n}$$

which is the usual pq/n form.

In conclusion we can say that the minimum-variance estimate of a parameter is the unbiased estimate that is based on the sufficient statistic when such exists. And the maximum likelihood technique finds it! In any case the M. L. E. has minimum variance.

I. Shortest Average Confidence Limits for Large Samples.

Though we have been concentrating on point estimation, it is proper to discuss this aspect of confidence interval estimation theory here since our friend $S(x_i; \theta)$ plays a role in it.

It seems natural enough to want our confidence interval for a population parameter as short as possible in some average sense. Generally this cannot be arranged except in large samples for certain population distributions. Rather than state for such cases a general theorem whose proof is more complicated than we care to discuss in these lectures, let us simply illustrate by taking the simple Bernoulli distribution

$$f_B(x; p) = p^x(1 - p)^{1-x} \qquad x = 0, 1.$$

Suppose we want 100c% confidence limits for p from a sample $(y_1, y_2, \cdots, y_n)$. Then

$$\int_{y_\ell}^{y_u} f_B(x; p)dx = \text{desired probability} = f(p).$$

We write the integral here to be general though our example would call for a Σ. To get the maximum probability for an interval is to get the

smallest interval for a fixed probability. So let's differentiate our last integral with respect to p and set the result equal to zero, viz.,

$$\frac{\partial}{\partial p} \int f(x;\, p)dx = \int \frac{\partial}{\partial p} f(x;\, p)dx = 0.$$

Since we want an expectation (average) we want our integral to have the factor $f(x;\, p)$. Hence we rewrite it

$$\int \frac{\partial}{\partial p} f(x;\, p)dx = \int \frac{\partial}{\partial p} [\log f(x;\, p)]\; f(x;\, p)dx = 0.$$

Another way of introducing the importance of the log here is as follows

1. $(y_1, y_2, \cdots, y_n)$ is a random sample,

2. $f(y_1;\, p) \times f(y_2;\, p) \times \cdots \times f(y_n;\, p)$ is its probability,

3. To maximize it (a minimum is obviously an end condition) we set the derivative with respect to p equal to zero, viz.,

$$\frac{\partial}{\partial p}[f(y_1;\, p) \times f(y_2;\, p) \times \cdots \times f(y_n;\, p)] = 0$$

and to get this into expectation form we write it

$$\left(\sum_{i=1}^{n} \frac{\partial \log f(y_i;\, p)}{\partial p}\right)[f(y_1;\, p) \times f(y_2;\, p) \times \cdots \times f(y_n;\, p)] = 0.$$

This requires the parenthetical sum to be zero.

Now to go back to our original plan and go on from there we need the additional fact mentioned earlier, namely,

$$Q = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(y_i;\, p)}{\partial p}$$

is approximately normally distributed with zero mean and with variance

$$\sigma_Q^2 = \frac{1}{n} E\left[\left(\frac{\partial \log f(y_i; p)}{\partial p}\right)^2\right].$$

Hence approximate 100c% confidence intervals are obtained by setting

$$\frac{Q}{\sigma_Q} = \pm Z_c$$

and solving for p. This interval is smallest.

Now for our Bernoulli distribution we follow this through.

$$\frac{\partial \log f(x; p)}{\partial p} = \frac{\partial}{\partial p}[x \log p + (1 - x)\log(1 - p)]$$

$$= \frac{x}{p} - \frac{1 - x}{1 - p}.$$

Next

$$E\left[\left(\frac{\partial \log f(x; p)}{\partial p}\right)^2\right] = E\left[\frac{x}{p} - \frac{1 - x}{1 - p}\right]^2$$

$$= \sum_{x=0}^{1}\left[\frac{x}{p} - \frac{1 - x}{1 - p}\right]^2 p^x (1 - p)^{1-x} = \frac{1}{p(1 - p)}.$$

Therefore

$$Q = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{p} - \frac{1 - y_i}{1 - p}\right] = \frac{\hat{p} - p}{p(1 - p)}$$

where

$$\hat{p} = (\Sigma y_i)/n.$$

Since $\sigma_Q^2 = \frac{1}{p(1 - p)} \times \frac{1}{n}$, we have

$$\frac{Q}{\sigma_Q} = \frac{(\hat{p} - p)}{\sqrt{\frac{p(1 - p)}{n}}}$$

and the 100c% confidence limits are obtained from solving

$$\frac{(\hat{p} - p)\sqrt{n}}{\sqrt{p(1 - p)}} = \pm Z_c$$

which is exactly the same result as that given on page 13.

VI. ORDER STATISTICS

When one tests the life of a sample of n items, it is obvious that if t_i denotes the time when the i-th one fails, then the data occur in a way that their serial order also gives them in order of increasing magnitude, i.e., $t_1 \leq t_2 \leq \cdots \leq t_n$. We say that the sample values are ordered by size in time. However, not all samples' values have this property so we consider rearranging the values in the sample $(x_1, x_2, \cdots, x_n)$ in increasing order of magnitude and then denoting this array by

$$(x_{(1)}, x_{(2)}, \cdots, x_{(n)}).$$

Consider all samples of size n from a base population. Then the smallest value in a sample varies randomly from sample to sample. So does the next-to-smallest, etc. Hence we have n new random variables each of which is called an <u>order statistic</u> as they are functions of a sample. We say $x_{(1)}$ is the first order statistic while $x_{(k)}$ is called the k-th order statistic. Now remember

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

and so these new random variables are not independently related. They are dependent in the strongest sense, namely, pairwise.

A. Typical Order Statistic Distribution.

Let x denote a random variable with continuous density function $f(x)$, $-\infty < x < \infty$ and for $n = 5$ let our random sample be $(x_1, x_2, \cdots, x_5)$. Consider, say, the fourth order statistic, $x_{(4)}$. Now for a particular

sample, $x_{(4)}$ might be any one of the five serially-ordered sample values, and, moreover, it can be any value in the domain of the random variable x. Suppose we say it is a particular value and denote this value by $x'_{(4)}$. Then what is the probability that the fourth order statistic will have a value in the interval

$$(x'_{(4)},\ x'_{(4)} + \Delta x_{(4)})?$$

More generally, let A be the event that a sample value lies in the interval $(-\infty,\ x'_{(4)})$, B be the event that a sample value lies in the interval $(x'_{(4)},\ x'_{(4)} + \Delta x_{(4)})$, and C be the event that a sample value lies in the interval $(x'_{(4)} + \Delta x_{(4)},\ \infty)$.

Now we ask how many equally likely samples satisfy the compound event A and A and A and B and C, whose probabilities, suggested by Figure 21,

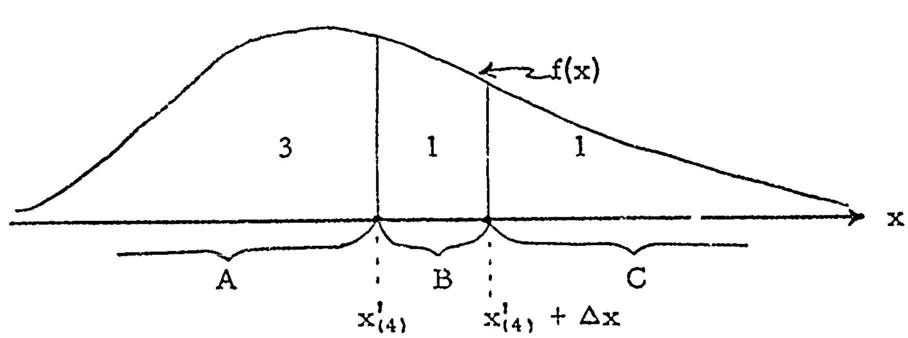


Figure 21

are

$$\Pr\{A\} = \int_{-\infty}^{x'_{(4)}} f(x)dx = F(x'_{(4)}),$$

$$\Pr\{B\} = \int_{x'_{(4)}}^{x'_{(4)} + \Delta x} f(x)dx = f(x_{(4)} + \theta\Delta x) \times \Delta x, \quad 0 < \theta < 1$$

$$\Pr\{C\} = \int_{x'_{(4)} + \Delta x}^{\infty} f(x)dx = 1 - F(x'_{(4)} + \Delta x).$$

Since eventually $\Delta x_{(4)} = \Delta x$ will go to zero, we may as well assume now that it is small enough to assure us that $x_{(5)}$ is greater than or equal to $x_{(4)} + \Delta x$. Then we can say that for any random sample $(x_1, x_2, \cdots, x_5)$ event A occurs three times since three of our five observations must be less than the fourth order statistic, event B the fourth order statistic's range occurs once as does event C, the fifth order statistic's range. We can indicate this by putting the numbers 3, 1, 1 in the three regions as shown in Figure 21. Hence such a typical sample would give the compound event of 3A's, 1B, and 1C which in turn has the probability

$$[\Pr\{A\}]^3 [\Pr\{B\}]^1 [\Pr\{C\}]^1 .$$

Once again we ask how many equally likely serially-ordered samples for each fixed set of five numerically ordered values would give this same situation. Well, let's first suppose $x'_{(4)} = x_1$, i. e., that the fourth smallest observation is the first observation.

Table XV lists the various different serial-numberings of these values which satisfy our requirement.

So there are four equally likely but different serial-numberings for our set of five values that give x_1 to be $x'_{(4)}$. Similarly we would find four equally likely but different ones for each of x_2, x_3, x_4, x_5 to be $x'_{(4)}$.

111

Therefore we would have to multiply the probability of 3A's, 1B, and 1C by 20 to get the probability of $x_{(4)}$ being in the interval $(x'_{(4)}, x'_{(4)} + \Delta x)$. This can be written

$$\Pr\{x_{(4)} < x'_{(4)} + \Delta x\} - \Pr\{x'_{(4)} < x'_{(4)}\}$$
$$= 20[F(x'_{(4)})]^3 [f(x'_{(4)} + \theta\Delta x)]\Delta x[1 - F(x'_{(4)} + \Delta x)]^1, \quad 0 < \theta < 1.$$

Now divide both sides of the last equation by Δx, and then let $\Delta x \to 0$. By definition the left member becomes the density function for $x_{(4)}$. If we denote it by $g(x_{(4)})$ then we have, dropping the prime on $x'_{(4)}$,

$$g(x_{(4)}) = 20[F(x_{(4)})]^3 [1 - F(x_{(4)})]^1 f(x_{(4)})$$

for all values of $x_{(4)}$ for which x is defined.

The probability density function just derived is readily obtainable as a particular application of the multinomial distribution. Just as we derived the binomial distribution by asking a question in a Bernoulli Process [Volume I, pages 73-75] we can obtain the multinomial distribution by asking a similar question in a more general process.

Table XV

Less Than $x'_{(4)}$	$x'_{(4)}$	Greater Than $x'_{(4)}$
$\{x_2, x_3, x_4\}$	x_1	$\{x_5\}$
$\{x_3, x_4, x_5\}$	x_1	$\{x_2\}$
$\{x_2, x_4, x_5\}$	x_1	$\{x_3\}$
$\{x_2, x_3, x_5\}$	x_1	$\{x_4\}$

B. The Generalized Bernoulli Process.

Suppose we have a process with the following characteristics:

On a trial (in a sequence of trials) some one of k different events $E_1, E_2, \cdots, E_k$ occurs;

The probability p_i with which E_i may occur remains fixed trial after trial. Note that $p_1 + p_2 + \cdots + p_k = 1$;

The trials are independent (i.e., the result on a trial is not affected by the result on a previous trial).

1. Question. In n successive trials, what is the probability of n_1 occurrences of E_1, n_2 occurrences of E_2, $\cdots$, n_k occurrences of E_k? Note that $n_1 + n_2 + \cdots + n_k = n$.

a. Argument. Proceeding as we did in the argument for the binomial distribution, we note there is a multiple random variable or multiple real-valued function in this question and it counts the number of "successes" of each event E_i in an element of the sample space. The sample space consists of all n-tuple arrays made up of any number of each of the E_i, with the total of such numbers being n, viz.,

$$\boxed{\begin{array}{l} e_1 = (E_1, E_1, \cdots, E_1) \\ \quad \ddots \\ \qquad e_r = (E_1, E_5, E_2, E_1, \cdots) \\ \qquad\quad \ddots \\ \qquad\qquad e_{k^n} \end{array}}$$

Sample space of k^n elements.

where e_r denotes the r^{th} series of trials. Therefore we shall call $X(e_r)$ = "the number of E_1's in e_r, the number of E_2's in e_r, $\cdots$, the number of E_k's in e_r." Then $X(e_r)$ can be any set of k integers in an ordered array where in each position there can be any integer from 0 to n.

In order to describe further the process and the random variable, we introduce, as before, the probability distribution of the ordered arrays of numerical outputs of this random variable. By way of illustration we give in Table XVI the development for the situation when n = 3 and k = 3.

Incidentally we can generate by the multinomial theorem of algebra all the various probabilities by expanding

$$(p_1 + p_2 + p_3)^3$$

and we can get each one from the compact formula

$$f(n_1, n_2) = \frac{3!}{n_1!\, n_2!\, n_3!}\, p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

where

$$n_i = 0, 1, 2, 3 \qquad \text{and}$$

$$\sum_{i=1}^{3} n_i = 3 .$$

We write the probability function as involving only 2 of the n_i since only 2 of them are functionally independent. Further examination of the table will indicate the coefficient on the product of a particular set of powers n_1, n_2, n_3 of the probabilities p_1, p_2, p_3, respectively, is simply the number of permutations of three things taken three at a time when n_1

Table XVI

e_r	$X(e_r)$	$Pr\{e_r\}$	Nr of E_1, E_2, E_3 (n_1, n_2, n_3)	$Pr\{X(e_r) = (n_1, n_2, n_3)\}$
$\{E_1, E_1, E_1\}$	3, 0, 0	p_1^3	(3, 0, 0)	p_1^3
$\{E_2, E_2, E_2\}$	0, 3, 0	p_2^3	(0, 3, 0)	p_2^3
$\{E_3, E_3, E_3\}$	0, 0, 3	p_3^3	(0, 0, 3)	p_3^3
$\{E_1, E_1, E_2\}$	2, 1, 0	$p_1^2 p_2$		
$\{E_1, E_2, E_1\}$	2, 1, 0	$p_1^2 p_2$	(2, 1, 0)	$3p_1^2 p_2$
$\{E_2, E_1, E_1\}$	2, 1, 0	$p_1^2 p_2$		
$\{E_1, E_1, E_3\}$	2, 0, 1	$p_1^2 p_3$		
$\{E_1, E_3, E_1\}$	2, 0, 1	$p_1^2 p_3$	(2, 0, 1)	$3p_1^2 p_3$
$\{E_3, E_1, E_1\}$	2, 0, 1	$p_1^2 p_3$		
$\{E_2, E_2, E_1\}$	1, 2, 0	$p_1 p_2^2$		
$\{E_2, E_1, E_2\}$	1, 2, 0	$p_1 p_2^2$	(1, 2, 0)	$3p_1 p_2^2$
$\{E_1, E_2, E_2\}$	1, 2, 0	$p_1 p_2^2$		
$\{E_2, E_2, E_3\}$	0, 2, 1	$p_2^2 p_3$		
$\{E_2, E_3, E_2\}$	0, 2, 1	$p_2^2 p_3$	(0, 2, 1)	$3p_2^2 p_3$
$\{E_3, E_2, E_2\}$	0, 2, 1	$p_2^2 p_3$		
$\{E_3, E_3, E_1\}$	1, 0, 2	$p_1 p_3^2$		
$\{E_3, E_1, E_3\}$	1, 0, 2	$p_1 p_3^2$	(1, 0, 2)	$3p_1 p_3^2$
$\{E_1, E_3, E_3\}$	1, 0, 2	$p_1 p_3^2$		
$\{E_3, E_3, E_2\}$	0, 1, 2	$p_2 p_3^2$		
$\{E_3, E_2, E_3\}$	0, 1, 2	$p_2 p_3^2$	(0, 1, 2)	$3p_2 p_3^2$
$\{E_2, E_3, E_3\}$	0, 1, 2	$p_2 p_3^2$		
$\{E_1, E_2, E_3\}$	1, 1, 1	$p_1 p_2 p_3$		
$\{E_1, E_3, E_2\}$	1, 1, 1	$p_1 p_2 p_3$		
$\{E_2, E_1, E_3\}$	1, 1, 1	$p_1 p_2 p_3$	(1, 1, 1)	$6p_1 p_2 p_3$
$\{E_2, E_3, E_1\}$	1, 1, 1	$p_1 p_2 p_3$		
$\{E_3, E_1, E_2\}$	1, 1, 1	$p_1 p_2 p_3$		
$\{E_3, E_2, E_1\}$	1, 1, 1	$p_1 p_2 p_3$		

are of one type, n_2 of another type, and n_3 of a third type [Volume I, pages 47-50].

In general, for n trials our probability rules tell us that a compound event which has n_1, "E_1's," n_2 "E_2's," $\cdots$, n_k "E_k's" has probability

$$p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}.$$

Since there are $P(n; n_1, n_2, \cdots, n_k)$ ways of arranging this number of "E_1's," "E_2's," $\cdots$, "E_k's," we conclude

$$\Pr\{X(e_r) = (n_1, n_2, \cdots, n_k)\} = P(n; n_1, n_2, \cdots, n_k) \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

where each of n_i can be any one of the values from 0, 1, $\cdots$, n with $n_1 + n_2 + \cdots + n_k = n$.

C. Application of the Multinomial Distribution to the Derivation of the Distribution Function of an Order Statistic.

Let us now reconsider the Section A and the obtaining of the probability density function of $x_{(4)}$ from a sample of size $n = 5$. In picking x five times at random from a population with p. d. f. $f(x)$, we want the selection to be such that

Event	Description of Event	Probability of Event	Nr of Occurrences
E_1	$x \epsilon (-\infty, x_{(4)})$	$F(x_{(4)})$	$n_1 = 3$
E_2	$x \epsilon (x_{(4)}, x_{(4)} + \Delta x)$	$f(x_{(4)})dx_{(4)}$	$n_2 = 1$
E_3	$x \epsilon (x_{(4)}, \infty)$	$1 - F(x_{(4)})$	$n_3 = 1$

Substituting the corresponding probabilities in the multinomial probability distribution function we have

$$\frac{5!}{3!\,1!\,1!}[F(x_{(4)})]^3\,[f(x_{(4)})dx_{(4)}]^1\,[1 - F(x_{(4)})]^1$$

which gives as the coefficient of $dx_{(4)}$ the probability density of the fourth smallest or second largest observation as we found before in the first section of this chapter.

D. Derivation of the General Order Statistic.

Consider the k-th order statistic from a random sample of size n from a population whose probability density is f(x). Just as we did in the previous section for the fourth order statistic from a sample of size 5, we invoke the multinomial theorem and probability function for the three events as shown in Figure 22.

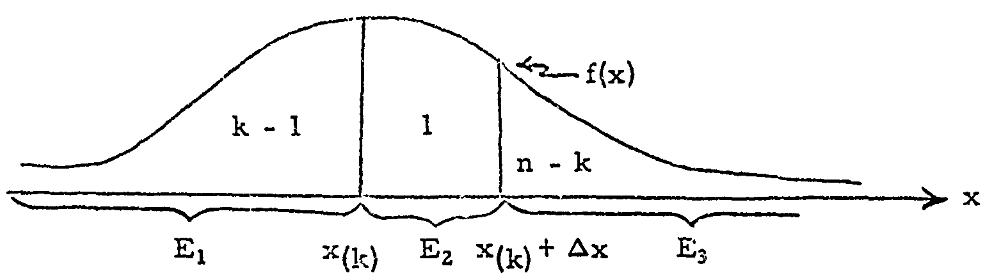


Figure 22

The areas of the three regions into which different numbers of observations of our size n sample fall are the three probabilities for the events E_1, E_2, E_3 as shown. Then event E_1 occurs (k - 1) times, E_2 once, and E_3 (n - k) times.

So it follows that if $g(x_{(k)})$ is the probability density function for $x_{(k)}$, that

$$g(x_{(k)})dx_{(k)} = \frac{n!}{(k-1)!\,(1)!\,(n-k)!} [F(x_{(k)})]^{k-1} [f(x_{(k)})dx_{(k)}]$$

$$\cdot [1 - F(x_{(k)})]^{n-k}$$

or

$$g(x_{(k)}) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} [F(x_{(k)})]^{k-1} [1 - F(x_{(k)})]^{n-k} f(x_{(k)}) .$$

It is interesting and useful to note that you can always write the cumulative distribution function $G(x_{(k)})$ of a single order statistic as an incomplete Beta function in terms of the cumulative function $F(x)$ of the random variable. You will recall that the Beta function $\beta(m, n)$ is defined in terms of the Gamma function by

$$\beta(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \int_0^1 y^{m-1}(1-y)^{n-1} dy$$

Now if you let $y = F(x)$, then we can write

$$G(x'_{(k)}) = \frac{1}{\beta(k,\ n-k+1)} \int_0^{y'} y^{k-1}(1-y)^{n-k} dy$$

where

$$y' = F(x'_{(k)}) .$$

By letting $k = 1$ we have the distribution function of the smallest element in the sample and when $k = n$ we obtain that for the largest. These are sometimes referred to as <u>Extreme Value Statistics</u>.

1. <u>Illustration</u>. Consider the sample $(x_1, x_2, \cdots, x_5)$ from $f(x) = 1,\ 0 < x < 1$. Then

$$g(x_{(1)})dx_{(1)} = \frac{5!}{0!\,1!\,4!}\left(\int_0^{x_{(1)}} 1\,dx\right)^0 (1\,dx_{(1)})^1 \left(\int_{x_{(1)}}^{1} 1\,dx\right)^4 .$$

This reduces to

$$g(x_{(1)}) = 5(1 - x_{(1)})^4, \qquad 0 \le x_{(1)} \le 1 .$$

Note that this distribution has a very high ordinate at $x_{(1)} = 0$ and drops off rapidly as $x_{(1)}$ increases, reaching 0 when $x_{(1)} = 1$. This is what you would get for a frequency distribution of the values of the smallest observation in repeated samples of size five.

On the other hand we get for the median, $x_{(3)}$,

$$h(x_{(3)}) = 30x_{(3)}^2(1 - x_{(3)})^2, \qquad 0 \le x_{(3)} \le 1$$

which is symmetric about $x_{(3)} = 1/2$ and has its highest value there, dropping off to zero as $x_{(3)}$ goes to zero or unity. These two order statistic distributions are shown in Figure 23.

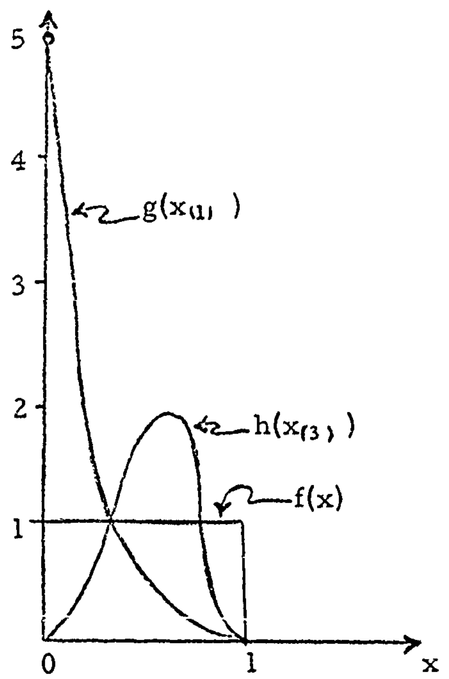


Figure 23

One would expect the smallest value to have greatest chances of being small while the median value would have little chance of being very small or very large. On the other hand the median value in repeated samples ought to be more frequently near the median of the population.

E. Maximum and Minimum Order Statistics.

The probability element of the minimum order statistic, $x_{(1)}$, from a random sample of size n from $f(x)$ is

$$n\left(\int_{x_{(1)}}^{\infty} f(t)dt\right)^{n-1} f(x_{(1)})dx_{(1)}$$

and of the maximum order statistic, $x_{(n)}$, is

$$n\left(\int_{-\infty}^{x_{(n)}} f(t)dt\right)^{n-1} f(x_{(n)})dx_{(n)} .$$

Next suppose $f(x)$ is uniformly distributed, $0 \le x \le 1$. Then we find

$$g(x_{(1)}) = n(1 - x_{(1)})^{n-1}, \quad 0 \le x_{(1)} \le 1.$$

Therefore the integral of $g(x_{(1)})$ from x to 1 is $(1 - x)^n$ which is the $\Pr\{x_{(1)} > x\}$. Therefore the cumulative distribution is

$$G(x) = \Pr\{x_{(1)} < x\} = 1 - (1 - x)^n, \quad 0 \le x \le 1.$$

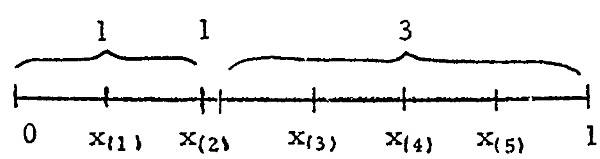
This is obvious from simpler considerations since it is the probability that not all n values of the sample fall into the interval $(x, 1)$. By elementary set reasoning we know this event is the complement of all values falling into the interval which has the probability $(1 - x)^n$.

Similarly the probability element of the n-th order statistic for the uniform distribution over $(0, 1)$ is

$$h(x_{(n)}) = nx_{(n)}^{n-1}, \quad 0 \le x_{(n)} \le 1.$$

Integration from 0 to x of $h(x_{(n)})$ gives x^n which is the $\Pr\{x_{(n)} < x\}$, the cumulative distribution $H(x)$. However we can get this directly for this simple base population since it is the probability that all the n values of the sample fall into the interval $(0, x)$.

1. Illustration. Let $f(x) = 2x$, $0 < x < 1$, and consider the random sample $(x_1, x_2, \cdots, x_5)$. Then the schematic diagram pictured here



suggests the probability density function

$$g(x_{(2)}) = \frac{5!}{1!\,1!\,3!}\left[\int_0^{x_{(2)}} f(x)dx\right]^1 \times \left[\int_{x_{(2)}}^{x_{(2)}+dx} f(x)dx\right] \times \left[\int_{x_{(2)}}^{1} f(x)dx\right]^3, \quad 0 \le x_{(2)} \le 1$$

$$= 40x_{(2)}^3(1 - x_{(2)}^2)^3 \qquad , \quad 0 \le x_{(2)} \le 1.$$

In a similar way we find

$$h(x_{(3)}) = 60x_{(3)}^5(1 - x_{(3)}^2)^2, \quad 0 \le x_{(3)} \le 1.$$

$$j(x_{(5)}) = 10x_{(5)}^9, \qquad 0 \le x_{(5)} \le 1.$$

F. Maximum Likelihood + Order Statistics.

1. Illustration. Suppose a random sample of size n is drawn from the exponential population with density function

$$f(x; \alpha, \beta) = \beta e^{-\beta(x-\alpha)}, \quad \alpha \le x < \infty, \quad 0 < \beta.$$

Then the likelihood of the compound event of the n sample values is

$$L = \beta^n e^{-\beta \sum_1^n (x_i - \alpha)},$$

and

$$\ln L = n \ln \beta - n\beta(\bar{x} - \alpha).$$

Taking the partial derivatives of $\ln L$ with respect to each parameter and setting each resulting expression equal to zero and then solving these two equations simultaneously gives us estimates $\hat{\alpha}$ and $\hat{\beta}$ which should locate a relative maximum for L, viz.,

$$\begin{cases} \dfrac{\partial \ln L}{\partial \alpha} = n\beta \\ \dfrac{\partial \ln L}{\partial \beta} = \dfrac{n}{\beta} - n(\bar{x} - \alpha) \end{cases}$$

and

$$\begin{cases} n\hat{\beta} = 0 \\ \dfrac{n}{\hat{\beta}} - n(\bar{x} - \hat{\alpha}) = 0 \end{cases}$$

Now the first equation gives $\hat{\beta} = 0$ which is not allowable in the second equation since it could not then give a finite $\hat{\alpha}$. So differentation so used fails as evidently no relative maximum exists!

Remember our purpose is to select an α and a β so that L is maximum. From the definition of L, we can see that for any $\beta > 0$, L would be greatest when $\hat{\alpha}$ takes on its largest possible value since then the exponent on e is largest. Now from the definition of f(x) we know that

all x are greater than or equal to α. Consequently α must be less than or equal to <u>every</u> value in the sample. Therefore the greatest value which $\hat{\alpha}$ can take on, consistent with the sample values, is the least value in the sample, i.e., $x_{(1)}$. So the maximum likelihood estimator $\hat{\alpha}$ for α is the first order statistic.

Next, substituting $x_{(1)}$ for $\hat{\alpha}$ in the second equation of the last pair of equations yields the maximum likelihood estimator for β,

$$\hat{\beta} = \frac{1}{\bar{x} - x_{(1)}} .$$

2. Illustration. For a random sample of size n from a population with uniform distribution $f(x) = 1/\theta$, $0 \le x \le \theta$, we have

$$L = \frac{1}{\theta^n}$$

$$\ln L = -n \ln \theta$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} .$$

Obviously we get nowhere setting this last expression equal to zero for this demands θ be infinite. Equally useless is to say that L is largest when θ is smallest and so let's take $\hat{\theta}$ to be $x_{(1)}$ the smallest value in our sample. Elementary considerations, on the other hand, tell us that $0 \le x_{(1)} \le \cdots \le x_{(n)} \le \theta$ and therefore we must have θ as big as $x_{(n)}$. Consequently, under this constraint, our maximum likelihood estimator is

$$\hat{\theta} = x_{(n)} .$$

Note the Cramer-Rao Inequality and the theory depending on it do not hold here since our parameter is a value at the end of the domain of the variable and hence not within an interval to permit using differentiation for relative minimum-maximum analysis. If one did not recognize this and calculated the lower bound for variance of the M. L. E. as given by the Cramer-Rao Inequality, he would get

$$\ell n\, f(x) = -\ell n\, \theta$$

$$\left(\frac{\partial\, \ell n\, f(x)}{\partial \theta}\right)^2 = \left(-\frac{1}{\theta}\right)^2 = \frac{1}{\theta^2}$$

$$nE\left\{\frac{\partial\, \ell n\, f(x)}{\partial \theta}\right\} = n \int_0^\theta \frac{1}{\theta^2} \times \frac{1}{\theta}\, dx = \frac{n}{\theta^2} .$$

Therefore

$$\text{var}\, \hat{\theta} \geq \theta^2/n$$

But $x_{(n)}$ is our M. L. E. and we can find its variance as follows:

$$g(x_{(n)})dx_{(n)} = n\left[\int_0^{x_{(n)}} \frac{1}{\theta}\, dx\right]^{n-1} \frac{dx_{(n)}}{\theta}$$

$$g(x_{(n)}) = \frac{n}{\theta}\,(x_{(n)}/\theta)^{n-1}$$

$$E\{x_{(n)}\} = \frac{n}{n+1}\,\theta, \quad E\{x_{(n)}^2\} = \frac{n}{n+2}\,\theta^2 .$$

Therefore

$$\text{var}\{x_{(n)}\} = \theta^2 \times \frac{n}{(n+1)^2(n+2)} ,$$

which is smaller than the lower bound to the variance found by using theory when the hypothesis for it was not satisfied. So all is well, if you look at all of it.

Incidentally the formula

$$E\{x_{(n)}\} = [n/(n+1)]\theta$$

is useful in predicting the maximum value in an assumed rectangular distribution when you have only a sample of size n. You simply use it in reverse and say $\hat{\theta}$ is $(n+1)/n$ multiplied by the sample maximum. To say when this is reliable and to what extent requires more analysis than we will go into here. One needs to calculate a probability statement about the difference between $[(n+1)/n]x_{(n)}$ and θ so as to get some sort of confidence interval for θ in terms of the $[(n+1)/n]x_{(n)}$ from a sample.

For n large, say $n \geq 100$, we know that the area under $g(x_{(n)})$ to the right of the mean, $[n/(n+1)]\theta$, is about .63 while in an interval of length $\theta/(n+1)$ to the left it is about .23. Hence about 85% of the time $x_{(n)}$ lies from $[(n-1)/(n+1)]\theta$ to θ. So we can say

$$\Pr\left\{-\frac{\theta}{n+1} < x_{(n)} - \frac{n}{n+1}\theta < \frac{\theta}{n+1}\right\} \doteq .86.$$

This can be written as

$$\Pr\left\{\frac{n}{n+1}\,\frac{n+1}{n}x_{(n)} < \theta < \frac{n}{n-1}\,\frac{n+1}{n}x_{(n)}\right\} \doteq .86.$$

So if you call $[(n+1)/n]x_{(n)}$ your estimate $\hat{\theta}$ of θ, then we can say that we are 86% confident that θ lies in

$$\left(\frac{n}{n+1}\hat{\theta}, \frac{n}{n-1}\hat{\theta}\right).$$

To see how effective this really can be suppose we assume a rectangular distribution and wish to estimate the upper bound or largest value from a sample of size 100. If the largest value in the sample is also 100, then our $\hat{\theta}$ is 101 and the 86% confidence interval is roughly (100, 103).

It must be remembered that the previous example was quite restrictive and that if the base population is other than rectangular, a new probability function for $x_{(n)}$ needs to be calculated, a new mean, and no doubt new considerations as to just what may be meant by a maximum in the base population, to say nothing as to the effect of the sample size n.

G. Confidence Interval + Order Statistic.

Suppose we assume the base population distribution of the previous illustration and then we ask for the smallest sample size such that we can be 99% certain that $x_{(n)}$ cuts off to the left the fraction β of the population. Well, this means we must find a sample size such that the following probability statement is true,

$$\Pr\{x_{(n)}/\theta > \beta\} = .99.$$

This can be evaluated as

$$1 - \int_0^{\beta\theta} g(x_{(n)})dx_{(n)} = 1 - \int_0^{\beta\theta} \frac{nx_{(n)}^{n-1}}{\theta^n}dx_{(n)}$$

$$= 1 - \left(\frac{x}{\theta}\right)^n \Bigg|_0^{\beta\theta}$$

$$= 1 - \beta^n = .99$$

Suppose we take $\beta = .95$. Then we have

$$\Pr\{x_{(n)}/\theta > .95\} = 1 - (.95)^n = .99$$

which gives

$$n = \frac{\ell n(.01)}{\ell n(.95)} \doteq 90.$$

This means if we take a sample of size 90, then we can be 99% confident that the largest value in our sample chops off at least 95% of the universe.

You might say we have a one-sided confidence interval for θ here since we can express this as

$$\Pr\{\theta < x_{(n)}/.95\} = .99 \quad \text{for } n = 90$$

or $(x_{(n)}, x_{(n)}/.95)$ is a 99% confidence interval for θ, the maximum value of the population.

If we take our previous illustration where $x_{(n)} = 100$, then we have 99% confidence in θ lying between 100 and 106 for about the same sample size. But if you look at the previous section you will note that the two-sided confidence interval there given really is

$$\left(x_{(n)}, \frac{n+1}{n-1} x_{(n)} \right)$$

a one-sided interval, and gives us less confidence, 85%, in a smaller interval, (100, 103), consistent with our later estimate.

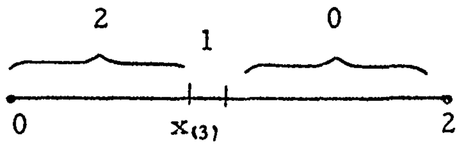
H. Hypothesis Testing with an Order Statistic.

1. Illustration. Our hypothesis tells us that a random variable x is distributed according to the law $f(x) = x/2$, $0 \le x \le 2$. We want to test this hypothesis by using for a test variate the value of the largest observation of three observations drawn at random from the base population. Using a one-sided critical region on the right with a level of significance of .05, let us determine whether the hypothesis should be accepted or rejected by the experiment which yielded the three sample values .211, 1.96, and 1.52.

a. Argument. The density function of $x_{(3)}$ is

$$g(x_{(3)}) = \frac{3!}{2!\,1!}\left[\int_0^{x_{(3)}} f(x)\,dx\right]^2 f(x_{(3)}), \quad 0 \le x_{(3)} \le 2$$

which is motivated by the diagram



Using the assumed form for $f(x)$, we find

$$g(x_{(3)}) = 3\left(\frac{x_{(3)}}{4}\right)^2 \times \frac{x_{(3)}}{2} = \frac{3}{32}\,x_{(3)}^5, \quad 0 \le x_{(3)} \le 2.$$

Now the largest value in the experimental sample is 1.96 which we must use, viz.,

$$\Pr\{x_{(3)} \geq 1.96\} = 1 - G(1.96)$$

$$= 1 - \int_0^{1.96} g(x_{(3)})dx_{(3)}$$

$$= 1 - \left[\frac{x_{(3)}^6}{64} \Bigg|_0^{1.96} \right] = 1 - (.98)^6$$

$$= 1 - .8865 = .1135 \nless .05$$

Therefore we do not reject the hypothesis.

VII. NONPARAMETRIC AND DISTRIBUTION-FREE TESTS

The two adjectives in the above title seem to be used alternatively and interchangeably in the literature. We will accept this, though argument can be given to distinguish between them.

Most of the tests we have used were based in some way on the assumption of normality. However, in practice we often know nothing about the parent population and so we need tests which do not depend on any assumption about the form of its distribution function. Distribution-free tests are based on order statistics or ordered samples, that is, we suppose the sample is ordered so that the observed data are arranged in increasing order of magnitude. In contrast to the common measures of location and dispersion, i.e., the mean and standard deviation with which we concern ourselves in parametric testing, here we use the median, quartiles, quantiles, etc., since they are sensitive to order by magnitude while the mean and standard deviation are not. In particular, when samples are small, distribution-free tests have proved safer than parametric ones where an error or lack of precise information concerning the required hypotheses has a rather dire consequence.

A. Sign Tests.

In earlier work we have tested, on the basis of a sample, whether a distribution was "located" at some prespecified point. Low let's test this nonparametrically. To do so we use the median $\tilde{x}$ of the sample to

estimate the true median $\tilde{\mu}$ of the base population. Suppose we wish to test whether some other number $\tilde{\mu}_0$ could be $\tilde{\mu}$.

Let $x_1, x_2, \cdots, x_n$ be our sample. Consider

Hypothesis: Median of distribution $= \tilde{\mu}_0$

Alternative: Median of distribution $\neq \tilde{\mu}_0$

To compute what is needed from our specific sample, we simply observe the signs of the differences

$$x_1 - \tilde{\mu}_0, \; x_2 - \tilde{\mu}_0, \cdots, x_n - \tilde{\mu}_0$$

and record the number of positive signs, y. Now $Y = y$ is a random variable since $(x_1, x_2, \cdots, x_n)$ is a random sample. Moreover $Y = y$ under this hypotheses has a binomial distribution.

$$f_B(y) = C_y^n/2^n, \qquad y = 0, 1, \cdots, n$$

since the probability of an observation falling to the right or left of the true median is 1/2 in either case. We might as well assume a continuous probability distribution for the base population X so that values equalling the median have probability zero and hence can be neglected.

So, in our sample, we find how probable is the particular value of Y and thereby make a decision about the $\tilde{\mu}_0$ which gave rise to it.

1. Illustration. For the sample demands 853, 857, 861, 851, 856, 859, 854, 849, consider the hypothesis that the median of the base population is 850, the alternative hypothesis being it isn't.

a. Argument. Apparently we should use a two-sided test, rejecting the hypothesis if y is either too large or too small. The test

statistic's probability distribution is

$$f_B(y) = C_y^8/256, \quad y = 0, 1, \cdots, 8.$$

which in tabular form is

y	0	1	2	3	4	5	6	7	8
f(y)	.004	.031	.109	.219	.274	.219	.109	.031	.004
F(y)	.004	.035	.144	.363	.637	.856	.965	.996	1.000

Before we use the particular value of y = 7 which our sample gives, we note from the above table that

$$\Pr\{y < 1 \text{ or } y > 7\} = \Pr\{y = 0 \text{ or } y = 8\}$$
$$= f(0) + f(8) = .008$$
$$\Pr\{y < 2 \text{ or } y > 6\} = \Pr\{y = 0 \text{ or } y = 1 \text{ or } y = 7 \text{ or } y = 8\}$$
$$= f(0) + f(1) + f(7) + f(8) = .070.$$

If we go back to our original derivation of the distribution function for the test statistic y and define y to count minus signs instead of plus signs, then the same value of $\tilde{\mu}_0$ would give us for the same sample set a value of $8 - y$ for y. This is why we here use a two-sided test to wash out the effect of this arbitrariness. In other words the rarity due to chance of a particular $\tilde{\mu}_0$ for candidacy for median must be considered so as to transcend this arbitrary choice in the definition of y. In our case this means we must think of $y = 1$ along with $y = 7$ as describing the actual situation of our data and hypothesized median.

Now if we decide to reject at the 1% level, then we see that y must be 0 or 8 to have probability less than 1% so that we would reject our

hypothesis. Hence in our particular situation where $y = 7$ our test accepts the hypothesis at the 1% rejection level. The same conslusion would maintain at the 5% rejection level since the actual case on hand occurs due to chance as seen through the eyes of our test statistic 7% of the time and is not rarer than 5% of the time. We can't get a total of 5% of probability from the tails of our discrete probability function.

If we lower our rejection level to 7%, then we would reject the hypothesis. But this is not a very stringent requirement for rejection.

In the same vein of thought but slightly more general lies the testing of whether two different samples $\{x_i\}$ and $\{y_i\}$ come from the same population. If the samples are fairly large, we can invoke the LaPlace-DeMoivre theorem as seen in the following case.

2. Illustration. Suppose we have two independent random samples $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ and that we wish to examine the possibility that they came from the same population with a distribution function which we do not know.

a. Argument. Now the x_i are not only random among themselves but also, under the assumption of a common base population distribution function, random among the y_i. Hence the probability is $1/2$ that any y_i is less than any x_i.

Let us prove this in general for any two independent random sample values y and x. If $f(x)$ is the common density function on $(0, \infty)$, then from the joint density function we get

/...

$$\Pr\{y < x\} = \int_0^\infty f(x)dx \int_0^x f(y)dy$$

since the admissible region in the x - y plane for our event "$y < x$" is as shown in Figure 24.

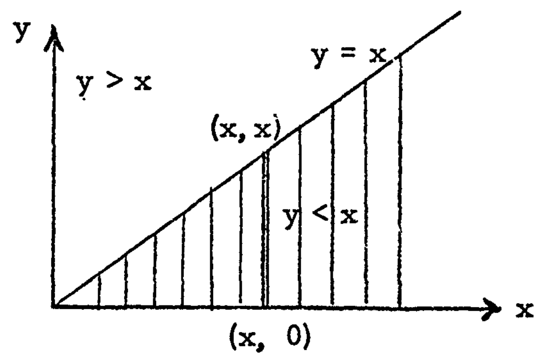


Figure 24

If we let $z = \int_0^x f(y)dy$, then $dz = f(x)dx$ and $z = 0$ when $x = 0$ while $z = 1$ when $x = \infty$. Therefore

$$\Pr\{y < x\} = \int_0^1 z dz = 1/2.$$

Thus we see our probability and the event are independent of $f(x)$ and hence distribution-free. So we are on firm ground to say for $z_i = x_i - y_i$ and for $u_i = 1$ if $z_i > 0$ and $u_i = 0$ if $z_i < 0$ that

$$\Pr\{u_i = 1\} = \Pr\{u_i = 0\} = 1/2.$$

Now u_i is a random variable and consequently so is $w = \Sigma u_i$. Its mean and variance are seen to be

$$E\{w\} = \Sigma E\{u_i\} = \left(\frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2}\right) = \frac{n}{2}$$

$$\text{var}\{w\} = \Sigma \text{var}\{u_i\} = \left(\frac{1}{4} + \frac{1}{4} + \cdots + \frac{1}{4}\right) = \frac{n}{4}.$$

Therefore the standard deviation of w is $\sqrt{n}/2$.

Recall how we proved the Central Limit Theorem in the previous course on pages 174-177. So if n is large enough, say over 30, we are pretty sure w, though discrete, can be adequately described by a normal distribution. This allows us to make the statement

$$\Pr\left\{-z_c < \frac{w - n/2}{\sqrt{n}/2} < z_c\right\} \doteq c$$

which leads us to the associated statement

$$\Pr\left\{\frac{n}{2} - z_c \frac{\sqrt{n}}{2} < w < \frac{n}{2} + z_c \frac{\sqrt{n}}{2}\right\} \doteq c.$$

To further exemplify how the large sample theory joins the distribution-free work, we take $c = .95$, $n = 100$, $n/2 = 50$, $\sqrt{n}/2 = 5$, $z_c = 1.96$, and find

$$\Pr\{50 - 1.96(5) < w < 50 + 1.96(5)\} \doteq .95$$

or

$$\Pr\{40 < w < 60\} \doteq .95.$$

We would therefore reject the hypothesis of the same population for the two sets of data each of size 100 at the 5% critical level if the sum of the u_i is greater than 60 or less than 40. Basically this is analogous to the first sign test; Σu_i is our binomially distributed test-variate with $p = 1/2$.

B. Point and Interval Estimation.

141

As we said earlier, the base population median is estimated by the sample median which is not unbiased but is consistent. Similarly we estimate the population quantiles by the corresponding sample quantiles. These are point estimators.

In contrast, to obtain a confidence interval estimate for $\tilde{\mu}$, we use the equal probability concept for an observation being to the right or left of $\tilde{\mu}$. It then follows that the probability that $x_{(r)}$, the r-th order statistic, exceeds $\tilde{\mu}$ is

$$\begin{aligned}
\Pr\{x_{(r)} > \tilde{\mu}\} &= \Pr\{x_{(i)} > \tilde{\mu},\ i = 1,\ 2,\ \cdots,\ n\} \\
&\quad + \Pr\{x_{(1)} < \tilde{\mu};\ x_{(i)} > \tilde{\mu},\ i = 2,\ 3,\ \cdots,\ n\} \\
&\quad + \Pr\{x_{(i)} < \tilde{\mu},\ i = 1,\ 2;\ x_{(j)} > \tilde{\mu},\ j = 3,\ 4,\ \cdots,\ n\} \\
&\quad - \\
&\quad + \Pr\{x_{(i)} < \tilde{\mu},\ i = 1,\ 2,\ \cdots, r - 1;\ x_{(j)} > \tilde{\mu}, \\
&\qquad\qquad j = r,\ \cdots,\ n\} \\
&= C_0^n\left(\frac{1}{2}\right)^n + C_1^n\left(\frac{1}{2}\right)^n + \cdots + C_{r-1}^n\left(\frac{1}{2}\right)^n.
\end{aligned}$$

So, if $f_B(i) = C_i^n\left(\frac{1}{2}\right)^n$, $i = 1,\ 2,\ \cdots,\ n$, then

$$\Pr\{x_{(r)} > \tilde{\mu}\} = \sum_{i=0}^{r-1} f_B(i)$$

Since $p = 1/2$ and hence our binomial distribution is symmetric, we also have

$$\Pr\{x_{(r)} > \tilde{\mu}\} = \sum_{i=n-r+1}^{n} f_B(i).$$

Further we find

$$\Pr\{x_{(s)} < \tilde{\mu}\} = \sum_{i=s}^{n} f_B(i) = \sum_{i=s}^{n} C_i^n / 2^n$$

and so

$$\Pr\{x_{(r)} < \tilde{\mu} < x_{(s)}\} = \sum_{i=r}^{s-1} f_B(i) = \sum_{i=r}^{s-1} C_i^n / 2^n, \quad r < s.$$

Thus $(x_{(r)}, x_{(s)})$ is a confidence interval for $\tilde{\mu}$ and the amount of confidence is the value of the sum of the probabilities in the right side of the last equation. These sums can be computed directly or by use of the tables of the Incomplete Beta function, e.g.,

$$1 - F_B(x) = \sum_{t=x+1}^{n} C_t^n p^t q^{n-t} = I_p(x+1, n-x)$$

$$= \frac{\int_0^p y^x (1-y)^{n-x-1} dy}{\int_0^1 y^x (1-y)^{n-x-1} dy} .$$

1. Illustration. For a sample of size 6, we find

a. $\Pr\{x_{(1)} < \tilde{\mu} < x_{(6)}\} = \frac{62}{64} = .97.$

b. $\Pr\{x_{(2)} < \tilde{\mu} < x_{(5)}\} = \frac{50}{64} = .78.$

2. Illustration. Suppose Q_1 is the lower quartile or .25 quantile. Then the probability of a random value being to the left of it is 1/4, to the right 3/4. Hence for a sample of size 6

$$\Pr\{x_{(1)} < Q_1 < x_{(4)}\} = \Pr\{x_{(1)} < Q_1 < x_{(2)}\}$$

$$+ \Pr\{x_{(2)} < Q_1 < x_{(3)}\}$$

$$+ \Pr\{x_{(3)} < Q_1 < x_{(4)}\}$$

$$= C_1^6\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^5 + C_2^6\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^4 + C_3^6\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^3$$

$$= \frac{1458 + 1215 + 540}{4096}$$

$$= \frac{3213}{4096} = .78$$

We note since the above sum is of three consecutive terms of the binomial distribution

$$f_B(i) = C_i^n\left(\frac{1}{4}\right)^i\left(\frac{3}{4}\right)^{n-i}$$

that this probability (confidence) could have been obtained from the Incomplete Beta as

$$I_{1/4}(1, 6) - I_{1/4}(4, 3).$$

We will talk about quantiles in general shortly.

C. Tolerance Limits.

You will recall we spoke of these earlier in Chapter III when we were estimating what size of spread would contain a certain percentage of the base population whose form of distribution was known. At that time we said we would return to the same concept when we no longer knew the base population distribution function. Let us begin by studying an example.

1. Illustration. Consider a random sample (x_1, x_2, x_3, x_4) and the random interval $[x_{(1)}, x_{(4)}]$. What proportion of future sampled items will fall in this region?

a. Argument. First we must recognize we can give only a qualified answer. That is, the proportion P we get will depend on the desired confidence we wish to put into it and vice versa. Now let us ask for 95% coverage. Then

$$\Pr\{P([x_{(1)}, x_{(4)}]) \geq .95\} = \frac{4!}{2!\,1!} \int_{.95}^{1} x^2 (1 - x)^1 \, dx$$

$$= 1 - I_{.95}(3, 2) = .015 .$$

Thus there is only a small chance that the interval $[x_{(1)}, x_{(4)}]$ will contain 95% of the probability of the distribution. Incidentally we used in the above equation the general formula for the probability element of the range V of a sample of size n, namely,

$$n(n - 1)V^{n-2}(1 - V)dV.$$

In general for a coverage $P \geq \beta$ and a confidence c we must solve

$$\int_{\beta}^{1} n(n - 1)V^{n-2}(1 - V)dV = 1 - (n\beta^{n-1} - (n - 1)\beta^n) = c$$

where

$$F_V(\beta) = \Pr\{V \leq \beta\} = 1 - c$$

which are usually intractable. Transcendental equations like this are hard to solve, usually solved by trial and error. In the case when β and

c are given and we wish to determine the smallest n for this desired tolerance interval, we must solve

$$F(\beta) = n\beta^{n-1} - (n-1)\beta^{n} = 1 - c$$

for n. An approximation is

$$n \doteq \frac{1}{4}\chi^{2}_{ch,4}\left(\frac{1+\beta}{1-\beta}\right) + \frac{1}{2}.$$

For example, when $c = .95$ and $\beta = .99$, we get

$$n \doteq \frac{1}{4}(9.488)\left(\frac{1.99}{.01}\right) + \frac{1}{2} = 473.$$

It is no wonder our original sample of size 4 gave us such a small chance of containing 95% of the probability of the distribution. As a matter of fact we need $n = 132$ to get 99% confidence that $[x_{(1)}, x_{(n)}]$ will account for 95% of the action.

This concept of working with a percentage of the probability and not with the same percentage of the range was first given by the late S. S. Wilks in two short classical papers. They mark at a later date as great a contribution as the earlier confidence interval did for a parameter of a distribution. It was known by Wilks and others that percentage of range could not be handled.

D. Confidence Intervals for Quantiles.

We call as usual x_p the p-th quantile point of a continuous cumulative distribution function $F(x)$ if $F(x_p) = p$.

Now $(x_{(k_1)}, x_{(k_1+k_2)})$ is a confidence interval for x_p having confidence

147

$$I_p(k_1, n+1-k_1) - I_p(k_1+k_2, n+1-k_1-k_2)$$

which is the

$$\Pr\{x_{(k_1)} < x_p < x_{(k_1+k_2)}\}$$

which in turn is really the probability of

$$F(x_{(k_1)}) < p < F(x_{(k_1+k_2)}).$$

Wilks tied up the essentials of all this in a very important theorem:

Wilk's General Theorem. If V_r = the sum of any r coverages of U's where $U_i = F(x_{(i)}) - F(x_{(i-1)})$, then the probability element of V_r is

$$\frac{n!}{(n-r)!\,(r-1)!} V_r^{r-1}(1-V_r)^{n-r} dV_r, \qquad 0 < V_r < 1$$

which is the Beta distribution for r and $n - r + 1$. The corollary of Wilk's General Theorem is also very useful and may be stated as:

Corollary. The average amount of probability for any one coverage is, taking $r = 1$,

$$E\{V_1\} = \frac{1}{\beta(n, 1)} \int_0^1 V_1(1-V_1)^{n-1} dV_1 = \frac{\beta(n, 2)}{\beta(n, 1)} = \frac{1}{n+1}.$$

It is no wonder some people say that confidence intervals on quantiles are equivalent to tolerance statements about the population with the same confidence.

E. Probability Paper Again.

The corollary just given by rights ought to be stated as:

Theorem. For any continuous distribution the expected values of the n + 1 probability areas determined by the random sample of n

values are all equal to each other and so their common value is $1/(n+1)$.

Remember how in Volume I on pages 108-111 we used arithmetic probability paper to check on normality in large samples. Now suppose we have a small sample, say size 10. Then as we promised on page 109 in Volume I, we would use order statistics plus the above theorem in order to give us a similar check with the same graph paper.

The practical value of this lies in the fact that for any random sample of size n, the total expected probability area to the left of the i-th order statistic is equal to $i/(n+1)$. Now if we took the points

$$\left(x_{(1)}, \frac{1}{n}\right), \left(x_{(2)}, \frac{2}{n}\right), \cdots, \left(x_{(n)}, \frac{n}{n}\right)$$

we could not plot the last one as it does not appear on our paper. The symmetry of the normal distribution suggests that whatever probability be assigned to $x_{(1)}$, then one minus it should be assigned to $x_{(n)}$. We could use the " spacings"

$$\left(x_{(1)}, \frac{1}{n+1}\right), \left(x_{(2)}, \frac{2}{n+1}\right), \cdots, \left(x_{(n)}, \frac{n}{n+1}\right)$$

or the " spacings"

$$\left(x_{(1)}, \frac{1}{2n}\right), \left(x_{(2)}, \frac{3}{2n}\right), \cdots, \left(x_{(n)}, \frac{2n-1}{2n}\right).$$

Much depends on your purpose in plotting. If you wish to obtain " optimum" estimates of the base population mean and standard deviation, you will find the literature replete with intricate analysis for each sample size. The last " spacings" given above are called by many authors

"intuitively plausible" and are found to be nearly as efficient as the optimum probability "spacings." Moreover they follow a simple formula. So we will use this spacing for plotting the order statistic cumulative probabilities on the "linearized" probability scale against the observed values of the sample which are measured on the arithmetic scale.

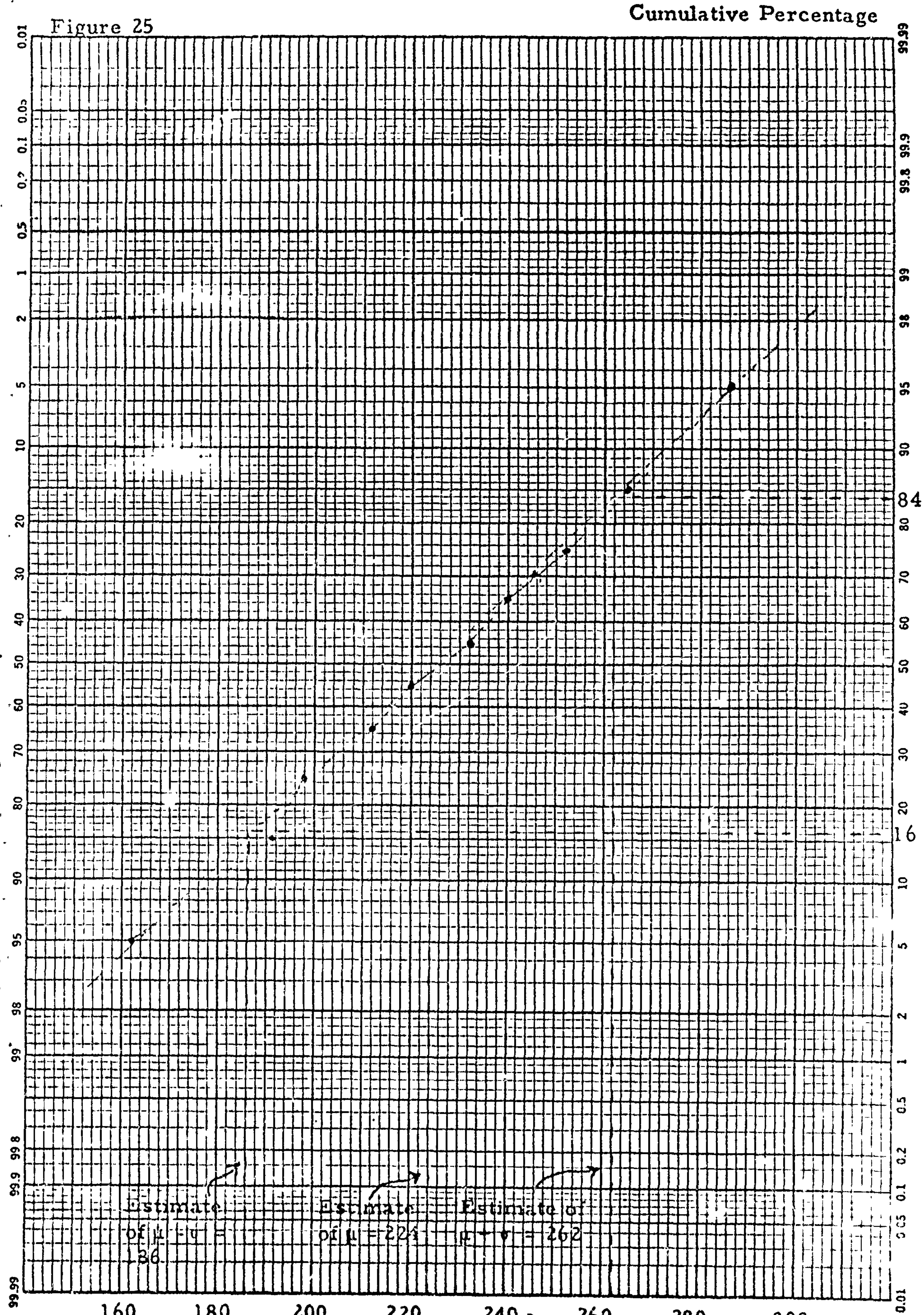
1. Illustration. The following ten demands were obtained randomly and then reordered: 162, 191, 198, 212, 220, 232, 240, 252, 265, 286. The corresponding cumulative probabilities for the associated order statistics' values, using $(2i - 1)/2n$ are

$$1/20,\ 3/20,\ 5/20,\ \cdots,\ 17/20,\ 19/20, \text{ respectively.}$$

On page 141 we see the plot or graph of these on arithmetic probability paper. They seem to lie near the straight line we drew in by eye so we accept normality of the base distribution. As before in large samples, we now estimate the mean by the 50th percentile which is 224 while 262 at the 84th percentile yields the estimate of 38 for σ, the standard deviation. The sample itself has a mean of 225.8 and a standard deviation of 37.1. Note if we took our old formula for estimating σ using the 16th percentile value we get $(262 - 186)/2 = 38$.

The same procedure can be used to estimate parameters for other type distributions on their own "linearized" probability paper. The spacings of the associated probabilities would change accordingly.

Figure 25


Cumulative Percentage
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01
84
16
Estimate of μ - σ = 186
Estimate of μ = 224
Estimate of μ + σ = 262
160 180 200 220 240 260 280 300

F. The Magnitude Test.

Suppose we have two random samples, $\{x_i\}$ and $\{y_i\}$, of demands and we wish to decide whether or not they come from the same parent population. Now we have seen how the sign test, when used in such a situation, considers only the signs of the differences $z_i = x_i - y_i$ and does not take into account the magnitude of these differences. Consider the following data from two samples of size 6,

i	1	2	3	4	5	6
z_i	3	4	-1	6	5	1

Under the assumption that the parent population is the same, it follows that the median of z_i is zero. Further it follows that the two x_i and y_i that give a value for z_i might just as well have been interchanged. This means any z_i could just as well have been positive or negative. So we might consider drawing a random sample of size 6 from the synthetic population of six possible pairs of differences, one drawing from each pair. This means each random sample uses a z_i with one sign. Hence our population consists of 64 equally likely possibilities, ranging from the extreme negative total of -20 when all signs are negative to that of +20 when all signs are positive. Our test variate is the sum of the six differences. The following table gives the frequency of occurrence of the various sums.

s	±20	±18	±16	±14	±12	±10	±8	±6	±4	±2	0
$f(s)$	1	2	1	1	3	4	4	4	4	5	6

So we see the probability of $s = 20$ is 1/64, of $s = -20$ is 1/64, of 18 is 2/64, of -18 is 2/64, etc. Obviously the distribution is symmetric about $s = 0$.

Now suppose we use a significance level of 5% and a one-sided test. Then the nearest we can come to it is by taking $s = 18$ which yields $Pr\{s \geq 18\} = 3/64 = .047$. The actual sample value is $s = 18$ which falls in the upper 5% critical region. Therefore, we would reject the null hypothesis which in this case is that zero is the median which in turn rejects the hypothesis that the two random samples came from the same distribution.

Recall the sign test takes no account of the magnitude. In this last example had we invoked the sign test, then the test variate would be $x =$ the number of positive signs and would have had the distribution

$$f_B(x) = C_x^6/2^6, \qquad x = 0, 1, 2, \cdots, 6$$

For the same significance level and right-sided test we find $x = 6$ is the only value falling in the critical region since $f(6) = .016$ while $f(5) + f(6) = .094 + .016 = .11$. Since $x = 5$ in the actual sample we would accept the null hypothesis that the number of positive signs equals the number of negative signs and hence that both samples come from the same distribution.

In a sense the magnitude test generalizes the sign test in that the former can be reduced to the latter by taking all possible arrangements of a fixed number of excesses and lumping the probabilities of their sums.

It is important to note from the previous discussion that for the same sample(s) we have come up with opposite decisions from two different hypotheses of randomness and their test variates. The lesson to be learned is that we usually solve an interpretation of a problem and not the problem per se.

G. Conditional Events.

Many practical situations call for an estimate of what to expect next after a sample has been taken. From another point of view we could suggest certain possibilities and then calculate their chances of occurring. It is this line of thought along which we will proceed. First let us look at

1. Illustration. Suppose we have had ten demands randomly given from a hypothetical distribution $f(x)$. For convenience we will assume $f(x)$ is defined over 0 to ∞. What we do will not be limited in application by this as the same result would evolve for a finite range of x. Now suppose three more demands are randomly drawn from $f(x)$. What is the probability that all three of these demands will be larger than any of the first sample?

a. Argument. We know that

$$g(x_{(10)}) = \frac{10!}{9!}\left[\int_0^{x_{(10)}} f(x)dx\right]^9 f(x_{(10)})$$

and the next three must be larger than $x_{(10)}$. The conditional probability that this happens is

$$\left[\int_{x_{(10)}}^{\infty} f(x)dx\right]^3 .$$

Therefore the joint probability function of these two events is

$$g(x_{(10)})\left[\int_{x_{(10)}}^{\infty} f(x)dx\right]^3 , \quad 0 \le x_{(10)} < \infty$$

Hence the probability of this event is

$$\int_0^\infty g(x_{(10)})dx_{(10)}\left[\int_{x_{(10)}}^\infty f(x)dx\right]^3$$

$$= 10\int_0^\infty \left[\int_0^{x_{(10)}} f(x)dx\right]^9\left[\int_{x_{(10)}}^\infty f(x)dx\right]^3 f(x_{(10)})dx_{(10)} .$$

If we let $U = \int_0^{x_{(10)}} f(x)dx$, then $dU = f(x_{(10)})dx_{(10)}$ and the integral becomes

$$10\int_0^1 (1 - U)^3\, U^9\, dU = 10\beta(4, 10) = 1/286$$

So we see the probability of the event of interest does not depend on the form of $f(x)$.

2. Illustration. For an arbitrary $f(x)$, $0 \le x < \infty$, find the probability that after a random sample of size n is drawn, the next two observations will lie outside the range of the sample.

a. Argument. This means that the $(n + 1)$st and $(n + 2)$nd demands lie outside of $x_{(1)} \le x \le x_{(n)}$. As we did in the previous illustration, we find the probability P of this event is given by

$$n(n - 1)\int_0^\infty f(x_{(1)})dx_{(1)}\int_{x_{(1)}}^\infty f(x_{(n)})\left[\int_{x_{(1)}}^{x_{(n)}} f(x)dx\right]^{n-2}$$

$$\times \left[1 - \int_{x_{(1)}}^{x_{(n)}} f(x)dx\right]^2 dx_{(n)}$$

Transforming by $U = \int_0^{x_{(1)}} f(x)dx$, $V = \int_0^{x_{(n)}} f(x)dx$, we see that $U \le V \le 1$,

$$V - U = \int_{x_{(1)}}^{x_{(n)}} f(x)dx$$

and so

$$P = n(n-1)\int_0^1 dU \int_U^1 (V-U)^{n-2}[1-(V-U)]^2 dV$$

$$= \frac{6}{(n+1)(n+2)}$$

free of the form of the distribution $f(x)$.

H. Elementary Protection Level Calculations.

The Theorem on page 138 can be used directly for a simple but typical protection level problem. Suppose we have a random sample of size n which is ranked from smallest to largest in our usual notation $x_{(1)}$ to $x_{(n)}$.

Now when we asked in the previous section about the probability of two or more additional random values behaving in a way conditioned to the original sample values, we ran into some calculus. However if we ask only about the next value, things are very simple. Suppose we ask for the probability that the next demand, call it x^*, is greater than, say $x_{(r)}$, $r \leq n$. Among the $n+1$ equally likely intervals created by the ordered values of the size n sample, we are asking for x^* to fall into any one of $n - r + 1$ of them. Therefore

$$\Pr\{x^* > x_{(r)}\} = \frac{n-r+1}{n+1}$$

151

To apply this suppose we want a protection level of .80 for an item whose demand is known over the past nine quarters. Then we want the probability of being out of stock to be no more than .20. Our formula says

$$r \geq 10 - .20(10) = 8.$$

This means that stocking up to the level of the eighth ranked previous demand reduces the probability of stockout to .20. This used the second highest previous demand.

To find that minimum sample size n for which x* need be only larger than the second highest previous value at different protection levels we offer the following table

Protection Level	.50	.60	.70	.80	.85	.90	.95
Minimum n	3	4	6	9	12	19	39

Related tables varying one or two of the three variables can be constructed to magnify this elementary concept of protection.

I. Tests of Randomness.

Since all of our previous theory and technique depended on the random selection or random occurrence of events or data, it sometimes is desirable to test selected data for this property. Actually the following tests are not capable of proving randomness exists if it does exist. They at best indicate to what degree nonrandomness exists. It is important to realize that each of these tests, even when they detect no nonrandomness, do not assure randomness. Some will frequently not detect nonrandomness when it is present. However, these tests are useful in avoiding faulty conclusions

because of incorrect assumptions and they can indicate the need for investigation of factors systematically affecting obtained results, that is, they can detect the presence of systematic variation.

Nonrandomness might be summed up by the following four characteristics of observed data:

1. discontinuities,
2. trends,
3. cyclic or periodic movement,
4. extreme values.

Bear in mind that the first three of the above characteristics are functions of the order in which the observed data, or the observed events from which we get the data, occur. Except for extreme values, nonrandomness as characterized by any of the other three symptoms usually can be made to disappear by a rearrangement.

1. Runs or sign test. Consider the following table giving three different orders of the same number of heads as of tails, each from 20 tosses of a coin.

Table XVII

Toss Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Order 1	T	T	T	T	H	H	T	T	H	T	H	T	T	H	T	H	T	T	H	T
Order 2	H	T	T	H	T	T	H	T	T	H	T	T	H	T	T	H	T	T	H	T
Order 3	T	T	T	T	T	T	T	H	H	H	H	H	H	H	T	T	T	T	T	T

The first series of heads and tails did occur randomly. The second and third ones are rearrangements of the first one. The first one does not appear unusual while the other two display some systematic effect. Now it can't be the number of heads since this is the same in each series. It is the order which signals our attention. The second series is made up entirely of sequences HTT. The third series is composed of a long run of T's followed by a long run of H's followed by a long run of T's.

To get at this in a way more scientifically revealing that which we noted above, let us examine such series for (1) length of a run of the same event, (2) number of runs of different lengths.

The first series has one run of length 4 (tails), two runs of length 2 (tails), two runs of length 2 (heads), five runs of length 1 (heads), three runs of length 1 (tails), or a total of thirteen runs. Note the shorter length runs occurred more frequently than the longer length runs. On the other hand the second series, obviously periodic, has more runs, fourteen, but only two different ones and these are of lengths 1 (heads) and 2 (tails). In the third series we have but three runs, each of great length. It appears that increasing the number of runs tends to reduce the length of runs and vice versa. Hence we seek a probabilistic description of both at once.

Actually what we did with the runs is abstractly equivalent to what we do with the signs of the differences of successive data. Hence it is a sophisticated form of the use of signs. For, if in place of H's and T's

we have twenty readings, then by taking the data in the order of its occurrences and simply using the signs as follows:

$$\text{sign}\,(x_2 - x_1),\ \text{sign}\,(x_3 - x_2),\ \cdots,\ \text{sign}\,(x_n - x_{n-1}),$$

we consider the runs in (+)'s and (-)'s.

In general, if we have two different entities, type A and type B, and if further we have in total m of type A and n of type B, then it can be shown that in a series of length m + n if U = number of runs from the m of type A plus the number from the n of type B,

$$\Pr\{U = 2\nu\} = \frac{2\binom{m-1}{\nu-1}\binom{n-1}{\nu-1}}{\binom{m+n}{m}}$$

$$\Pr\{U = 2\nu + 1\} = \frac{\binom{m-1}{\nu}\binom{n-1}{\nu-1} + \binom{m-1}{\nu-1}\binom{n-1}{\nu}}{\binom{m+n}{m}}$$

when the series is random. Extreme values for U, small ones indicating a few long runs, large ones indicating many short runs, have low probability and hence indicate possible nonrandomness.

a. Illustration 1. In 11 successive quarters demands for a certain FSN appeared as follows

3, 7, 8, 10, 11, 13, 12, 11, 7, 6, 8.

We will assume they appeared randomly from a stable distribution. By taking successive differences we find the sequence of signs is

+ + + + + - - - - +

consisting of three runs. Now how probable due to chance is the case of

three or fewer runs in a sequence of 5 plus signs and 5 minus signs? Well, the probability of at most three runs is

$$\Pr\{U \le 3\} = \Pr\{U = 3\} + \Pr\{U = 2\}$$

since we cannot have less than two different runs. Hence, by taking $\nu = 1$ in each of our previous formulae along with $m = n = 5$, we find

$$\Pr\{U \le 3\} = \frac{2\binom{4}{0}\binom{4}{0}}{\binom{10}{5}} + \frac{\binom{4}{1}\binom{4}{0} + \binom{4}{0}\binom{4}{1}}{\binom{10}{5}}$$

$$= \frac{2}{252} + \frac{8}{252} = .0397$$

So this supposed random sequence of demands has a property that occurs due to chance only 4% of the time. When this small number of runs occurs, it is very likely that some nonrandom behavior is present.

We must remember, in order to use this method, to transform our data into a sequence of events of two kinds. Commonly one designates an element as above or below the median, thereby creating two classes. This has the advantage of always making equal the number of elements of each kind, i.e., we can always take $m = n$ in our previous formulae. If there are an odd number of elements, we drop the median. Miss Swed and Dr. Eisenhart have given a table for this case, that is, for $m = n$. A part of it follows in Table XVIII. The entries give the number of runs for a particular number of elements $2m = 2n$ such that the probability of this number of runs or less than (greater than) this number is α for $\alpha = .05$ and for $\alpha = .01$.

Table XVIII

Critical Values of U = Number of Runs

	Lower Critical		Upper Critical	
m = n	α = 0.05	α = 0.01	α = 0.05	α = 0.01
5	3	2	9	10
6	3	2	11	12
7	4	3	12	13
8	5	4	13	14
9	6	4	14	16
10	6	5	16	17
20	15	13	27	29

When m is large, theory tells us that we can use the fact the run distribution is nearly normal with expected value m + 1 and standard deviation $\sqrt{m/2}$.

2. <u>Mean square successive difference test.</u> This test is more powerful but not as quick and easy to apply as the tests in the former section. The former tests were distribution-free whereas this one is not. This test depends on a statistic whose distribution was discovered by von Neumann. As happened so often with our classical distribution functions, he assumed a normal distribution for his base population from which the random samples come.

Here we must compute the average of the squares of the (n - 1) successive differences between successive elements in a random sample

of size n. Now we can prove the expected value of this statistic, namely of

$$\delta^2 = \frac{\Sigma(x_{i+1} - x_i)^2}{n-1}$$

is $2\sigma_x^2$, regardless of the base population distribution. But the expected value of the ordinary sample variance, namely of

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

is σ_x^2. Therefore we can say the ratio

$$\eta = \frac{\delta^2}{s^2}$$

has expected value 2. Dr. von Neumann gave us the distribution function for η, and in 1942 Dr. Hart gave a table of its values. We repeat, as has happened in so many other situations, Dr. von Neumann assumed the sample came from a normal distribution. Table XIX is an abbreviated form of Dr. Hart's table.

Before we illustrate the use of η in detecting nonrandomness in a sample, we might get a feeling for its sensitivity to nonrandomness by noticing how it might vary from the value of 2 in certain situations. For example, when data has an upward trend, δ^2 will increase much less than s^2. So η would be less than 2. On the other hand, if the data rapidly goes up and down, δ^2 will increase proportionally greater than s^2. Then η will be greater than 2.

Table XIX

Critical Values for η

Sample Size n	Lower Critical $\alpha = 0.05$	Lower Critical $\alpha = 0.01$	Upper Critical $\alpha = 0.05$	Upper Critical $\alpha = 0.01$
4	0.78	0.63	3.22	3.37
5	0.82	0.54	3.18	3.46
6	0.89	0.56	3.11	3.44
7	0.94	0.61	3.06	3.39
8	0.98	0.66	3.02	3.34
9	1.02	0.71	2.98	3.29
10	1.06	0.75	2.94	3.25
20	1.30	1.04	2.70	2.96

a. Illustration 1. This illustration was first given by C. A. Bennett of General Electric. He wished to show that the runs test is not as powerful as the mean square successive differences test. He gave the following results of measuring a standard sample in the order of their analysis.

First let us compute η.

Table XX

Sample Nr i	Result x_i	Difference $x_{i+1} - x_i$	Sample Nr i	Result x_i	Difference $x_{i+1} - x_i$
1	83.50	0.13	11	84.40	0.10
2	83.63	0.53	12	84.50	0.38
3	84.16	-0.91	13	84.88	-0.34
4	83.25	0.11	14	84.54	0.16
5	83.36	0.90	15	84.70	0.10
6	84.26	-0.26	16	84.80	-0.56
7	84.00	0.61	17	84.24	-0.13
8	84.61	-0.15	18	84.11	0.41
9	84.46	-0.26	19	84.52	-0.38
10	84.20	0.20	20	84.14	

Now

$$\Sigma x_i = 1684.26 \qquad \Sigma x_i^2 = 141840.7214.$$

Hence

$$\Sigma(x_i - \overline{x})^2 = 4.1341.$$

For n = 20 we then find

$$s^2 = \frac{1}{n-1} \Sigma(x_i - \overline{x})^2 = 0.2176.$$

Next

$$\delta^2 = \frac{1}{n-1} \Sigma(x_{i+1} - x_i)^2 = \frac{3.4664}{19} = 0.1824$$

Therefore

$$\eta = \frac{0.1824}{0.2176} = 0.838.$$

Going back to Table XIX, we see that there are only two chances in 100 of η falling outside the interval (1.04, 2.96) when n = 20. Hence our computed value of η is significant of nonrandomness being present. An examination of the original data indicates an upward trend.

Now let us use our earlier sign or runs test on this data. There is a total of nine runs considering the runs above and below the median of 84.25. For n = 20 the expected number of runs is eleven and though nine is smaller, it is not significantly small at the 5% level which is six runs as can be seen from Table XVIII when m = n = 10. Therefore our runs test does not detect the nonrandomness which the other test does detect.

There are other tests based on runs--the length of the longest run, the distribution of runs, etc. No one test is best to detect nonrandomness in all cases. For example, the test based on the number of runs may not indicate nonrandomness while the test based on the longest run will.

APPENDIX A. THEORY AND PRACTICE

We have spoken of the relation between probability and relative frequency. It has been said that the gap between probability theory and practice is a difficult one to bridge. One bridge over the gap is "the law of averages," known in probability theory as the law of large numbers. We can speak of it here since it refers to the situation in which there is a sequence of independent events with fixed probability p. If a sequence of n trials is made and the number of successes is S_n, the proportion of successes in n trials is S_n/n. We ought to have some feeling that the average S_n/n approaches the fixed probability p as the number of trials gets larger and larger.

To this end let us consider the repeated tossing of an unbiased coin and keeping tract of the proportion of heads. The law of large numbers tells us that our hopes are not in vain; in some sense this proportion should approach 1/2. Now we shouldn't expect this proportion to suddenly become exactly 1/2. So let's take some small percentage of deviation, ϵ, and ask, for each number of trials n, what is the probability that the proportion of heads differs from 1/2 by less than ϵ. Specifically let $\epsilon = 10\%$. Then we have to find for each n the probability that the percentage of heads lies between 40% and 60%. Now we need a probability measure of the set of favorable sequences, i.e., those which will not deviate from 50% heads by more than 10%. Our Binomial Law provides us with this. It says the probability of exactly k heads in a sequence of n tosses is ${}_nC_k(.5)^k(.5)^{n-k}$.

A sequence is tolerable if the number of heads k satisfies

$$|k/n - .50| \leq .10$$

If $n = 5$, then tolerable tosses have 2 or 3 heads since the ratios $2/5 = .40$ and $3/5 = .60$ are within the tolerance limits - they are the limits. On the other hand a sequence with 0, 1, 4, or 5 heads is "out." The probability of a "tolerable" sequence is the sum of the terms in the Binomial Law for those values of k for which k/n is "within limits." In the case of $n = 5$ this says we must add the terms of the expansion $(.50 + .50)^5$ for which k is 2 or 3, that is,

$$10(.5)^2(.5)^3 + 10(.5)^3(.5)^2 = .63$$

is the probability of an acceptable sequence of 5 tosses.

If you were to go on with this by taking larger values for n, keeping $\epsilon = 10\%$, you would obtain among others the entries

n	Number of Heads Acceptable	Probability
5	2 or 3	.63
10	4, 5, or 6	.66
15	6, 7, 8, or 9	.70
20	8, 9, 10, 11, or 12	.74
100	40, $\cdots$, 60	.96
200	80, $\cdots$, 120	.996

So we see the probability of acceptable sequences which deviate from 50% heads by not more than 10% steadily increases as we toss the coin more and more. However, note that no matter how large n may

become the extreme cases of all heads or all tails and similar sequences are still possible. For example, when n = 200, they are included in the .004 fraction of the sequences, the undesirable ones. They just are less and less probable.

REFERENCES

Chapter I

1. Clopper, C. J. and Pearson, E. S., "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," Biometrika, Vol 26 (1934), 404-420.

2. LaPlace, P. S., "Théorie Analytique des Probabilités," third edition, Paris, Gauthier-Villars, 1886.

3. Neyman, J., "First Course in Probability and Statistics," New York: Henry Holt and Company, 1957.

4. Ricker, W. E., "The Concept of Confidence or Fiducial Limits Applied to the Poisson Frequency Distribution," Journal of the American Statistical Association, Vol 32 (1937), 349-356.

5. Thompson, C. M., "Tables of Percentage Points of the Chi-square Distribution," Biometrika, Vol 32 (1941), Part II, 187.

6. Wilks, S. S., "Elementary Statistical Analysis," Princeton, N. J.: Princeton University Press, 1949.

7. Wilks, S. S., "Mathematical Statistics," New York: John Wiley and Sons, 1962.

8. Wilson, E. B., "Probable Inference, The Law of Succession, and Statistical Inference," Journal of the American Statistical Association, Vol 27 (1927), 209-212.

Chapter II

1. Elderton, W. P., "Frequency Curves and Correlation," fourth edition, Washington D. C.: Harren Press, 1953.

2. Fisher, R. A., "On a Distribution Yielding the Error Functions of Several Well-known Statistics," Proceedings of the International Mathematics Congress, Toronto, Vol II (1924), 805-813.

3. Gossett, W. S., "The Probable Error of a Mean," Biometrika, Vol 6 (1908), 1-25.

4. Gossett, W. S., "New Tables for Testing the Significance of Observations," Metron, Vol 5 (1925), 105-112.

5. Pearson, K., "Contributions to the Mathematical Theory of Evolution," Philosophical Transactions, Vol 185 (1894), p. 71; Vol 186 (1895), p. 343; Vol 191 (1898), p. 229.

6. Pearson, K., "On a Criterion that a System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Risen in Random Sampling," Philosophical Magazine, Vol 50 (1900), 157-175.

7. Pearson, K., "On the Systematic Fitting of Curves to Observations and Measurements," Biometrika, Vol 1 (1902), p. 265; Vol 2 (1902), p. 1.

8. Pearson, K., "On the Curves Which are Most Suitable for Describing the Frequency of Random Samples of a Population," Biometrika (1905), Vol 5, 172-175.

9. Snedecor, G. W., "Statistical Methods," Ames, Iowa: Iowa State College Press, 1937.

10. Wilks, S. S., "Elementary Statistical Analysis," Princeton, N. J.: Princeton University Press, 1949.

11. Wilks, S. S., "Mathematical Statistics," New York: John Wiley and Sons, 1962.

Chapter III

1. American Society for Testing Materials, "Manual on Quality Control of Materials," Parts 1-3, Philadelphia, 1951.

2. Ehrenfeld, S. and Littauer, S., "Introduction to Statistical Method," New York: McGraw-Hill Book Company, 1964.

3. Natrella, M. G., "Experimental Statistics," Government Printing Office, Washington D. C., 1964.

4. Wilks, S. S., "On the Determination of Sample Sizes for Setting Tolerance Limits," Annals of Mathematical Statistics, Vol 12 (1941), 91-96.

5. Wilks, S. S., "Statistical Prediction with Special Reference to the Problem of Tolerance Limits," ibid, Vol 13 (1942), 400-409.

Chapter IV and V

1. Cramer, H., "Methods of Mathematical Statistics," Princeton, N. J.: Princeton University Press, 1946.

2. Fisher. R. A., "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error," Monthly Notices of the Royal Astronomical Society, Vol 80 (1920), 758-770.

3. Fisher, R. A., "Theory of Statistical Estimation," Proceedings of the Cambridge Philosophical Society, Vol 22 (1925), Part 5, 700-725.

4. Mills, H. F., ALRAND Report 24, "Program Sixty One," Ships Parts Control Center, Mechanicsburg, Pa., 1961.

5. Natrella, M. G., "The Relation Between Confidence Intervals and Tests of Significance," The American Statistician, Vol 14 (1960), 20-23.

6. Natrella, M. G., "Experimental Statistics," Washington D. C.: Government Printing Office, 1964.

7. Neyman, J. and Pearson, E. S., "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," Biometrika, Vol 20A (1928), Part 1, pp. 175-240; Part 2, pp. 263-294.

8. Neyman, J. and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," Transactions of the Royal Society of London, Series A, Vol 231 (1933), 289-337.

9. Statistical Research Group, Columbia University, "Sequential Analysis of Statistical Data: Applications," New York: Columbia University Press, 1945.

10. Statistical Research Group, Columbia University, "Techniques of Statistical Analysis," New York: McGraw-Hill Book Company, 1947.

11. Wilks, S. S., "Mathematical Statistics," New York: John Wiley and Sons, 1962.

<u>Chapter VI and VII</u>

(176)

1. Bradley, J. V., "Distribution-Free Statistical Tests," WADD Technical Report 60-61, Office of Technical Services, U. S. Department of Commerce, Washington D. C., 1960.

2. Chernoff, H. and Lieberman, G., "Use of Normal Probability Paper," Journal of the American Statistical Association, Vol 49 (1954), 778-784.

3. Fraser, D. A. S., "Nonparametric Methods in Statistics," New York: John Wiley and Sons, 1957.

4. Mosteller, F., "On Some Useful 'Inefficient' Statistics," Annals of Mathematical Statistics, Vol 17 (1946), 377-408.

5. Ogawa, J., "Contributions to the Theory of Systematic Statistics," Osaka Mathematical Journal, Vol 3 (1951), 175-213.

6. Sarhan, A. E. and Greenberg, B. G., "Contributions to Order Statistics," New York: John Wiley and Sons, 1962.

7. Siegel, S., "Nonparametric Statistics for the Behavorial Sciences," New York: McGraw-Hill Book Company, 1956.

8. Wilks, S. S., "Mathematical Statistics," New York: John Wiley and Sons, 1962.

9. Wilks, S. S., "Order Statistics," Bulletin of the American Mathematical Society, Vol 54 (1948), 6-50.